Modelling the Production and Propagation of Sound in Individual Human Vocal Tracts

by

Lukas Schickhofer

January 2019
Technical Reports
KTH Royal Institute of Technology
Department of Mechanics
SE-100 44 Stockholm, Sweden

©Lukas Schickhofer 2019
Universitetsservice US–AB, Stockholm 2019
Lukas Schickhofer  
Modelling the Production and Propagation of Sound in Individual Human Vocal Tracts  
Linnaeum Flow Centre  
Department of Mechanics, KTH Royal Institute of Technology  
SE-100 44 Stockholm, Sweden  

Abstract  
Voice generation and the expression through speech are of vital importance for communication. The human upper airways are the origin of the process of speech production, which involves a modulation of the periodically pulsed pressure from the lungs by the vocal tract volume. In this work, phonation and voiced speech are investigated through both low- and high-order models, which are applied to vocal tract geometries of increasing complexity. Initially, the effect of variations of vocal fold closure, fundamental frequency, and vocal tract length on the computed acoustic signal is examined through parameter studies based on one-dimensional wave reflection analogues. Eventually, unsteady large eddy simulations based on the compressible Navier-Stokes equations are carried out to compute the pressure fluctuations and the associated distribution of resonance modes as a result of the interaction with the static vocal tract. Thus it is possible to calculate tonalities from the entire audible range of frequencies from 20 to 20000 Hz. In particular the inharmonic broadband sound component produced predominantly by coherent structures in the upper airways and at frequencies above 2 kHz is resolved in the current study, which is not captured by low-order models based on wave equations. Furthermore, three-dimensional numerical meshes based on surface representations of the human upper airways under voiced speech from magnetic resonance imaging (MRI) data of a healthy male subject are applied. These are necessary to resolve high-order acoustic modes that would not be represented by simplified geometries. Validation and verification of the chosen methods are achieved through comparison with experimentally obtained speech data, as well as Helmholtz eigenfrequencies of the considered vowel pronunciations. The main scope of this work is the assessment of acoustic sources and the conditions for aerodynamic sound being produced and propagated in the upper airways during phonation. The distribution of acoustic sources involved in the generation of the dominant frequencies are identified by application of acoustic analogies as well as surface Fourier transformation of the acoustic pressure fluctuations.  

However, the human upper airways do not only embrace the source of phonation and affect the modulation of the voice. Moreover, unwanted sounds may be generated in the upper airways due to elastic, collapsible parts that are susceptible to flow-induced vibration and resonance. The sound resulting from fluid-structure interaction in the upper respiratory tract, commonly known as snoring, can be an important indicator for underlying breathing disorders,
such as obstructive sleep apnea (OSA). In a smaller part of this project, the flow structures and acoustic sources as a result of the interaction of shear flow of various Reynolds numbers with an elastic element are computed. The geometric dimensions are chosen to be representative of average physical values of the upper respiratory tract. Onset of tissue vibrations and resonance effects are investigated for a range of parameters of both solid and fluid.

The obtained results of this work are aimed to contribute also to the development of a computational tool that assists physicians in the assessment of the airway function and the effectiveness of treatment plans prior to their application.

**Key words:** Biomechanics, vocal tract acoustics, numerical flow simulation, fluid-structure interaction
Lukas Schickhofer
Modellering av Produktion och Spredning av Ljud i Människans Luftvägar

Linné Flow Centre
Department of Mechanics, KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden

Sammanfattning


De mänskliga övre luftvägarna utgör inte enbart källan till tal och moduleringen av detsamma. Därutöver kan önskat ljud genereras på grund av elastiska, ständigbara delar som är mottagliga för flödesinducerade vibrationer och resonans. Ljudet som resulterar från fluid-strukturinteraktionen i de övre luftvägarna, vanligen känt som snarkning, kan vara en viktig indikator för undantagande andningssvårigheter såsom obstruktiv sömnapné (OSA). I en mindre del av detta arbete beräknades flödesstrukturer och akustiska källor som uppträtt vid interaktion mellan skjuvflöden vid olika Reynolds tal och ett elastiskt
element. De geometriska dimensionerna valdes för att motsvara typiska fysiska värden hos de övre luftvägarna. Uppkomsten av vävnadsvibrationer och resonanseffekter undersöktes för ett spann av flödes- och strukturparameterar.

Resultaten från detta arbete syftar även till att bidra till utvecklingen av ett beräkningsverktyg för att understödja läkare i utvärdering av luftvägsfunktion och effektivitet av behandlingsplaner innan de sätts in.

Nyckelord: Biomekanik, luftvägsakustik, numeriska flödesberäkningar, fluid- strukturinteraktion
Conferences

Parts of this work have been presented at the following conferences. The presenting author is underlined.


L. Schickhofer, A. Dahlkild, M. Mihaescu. On direct aeroacoustics calculations of the vocal tract. 11th ERCOFTAC Workshop on Direct and Large-Eddy Simulation. Pisa (Italy), 2017

L. Schickhofer, M. Mihaescu. Flow structures and dominant frequencies in the vocal tract. 6th International Conference on Jets, Wakes and Separated Flows. Cincinnati (USA), 2017

L. Schickhofer, J. Malinen, M. Mihaescu. Flow instabilities in the upper airways during phonation. 8th World Congress of Biomechanics. Dublin (Ireland), 2018

L. Schickhofer, J. Malinen, M. Mihaescu. Flow-induced sound generation during voiced speech. BIT Circus - Numerical Mathematics and Computational Science. Helsinki (Finland), 2018

Preface

The underlying thesis focuses on modelling of the mechanisms of sound generation and propagation in the human upper airways under voice production. Methods for the computation and prediction of acoustic pressure waves arising from the flow inside the vocal tract are investigated and used on realistic models. A short chapter of the thesis is dedicated to the fluid-structure interaction under physiological conditions and the computation of acoustic sources in a simplified geometry.

The thesis is written in two parts, with the first part giving a summary of the research topic alongside selected results and the second part containing the added papers.

In the first chapter an introduction to the topic of this work is given by presenting the basics of the anatomy and physiology of the human upper airways, which represent the region of interest. Additionally, their main functions in our body and during phonation are discussed.

Chapter 2 focuses on the motivation for this project and highlights the most important questions to be answered. The current theory of speech production is presented. Additionally, the pathology of upper airway diseases, both with respect to phonation, as well as to the respiratory function is reviewed, in order to highlight the clinical relevance of the obtained results.

Chapter 3 gives details on the applied methods and models for the numerical study of voice production in this work.

Chapter 4 gives a brief overview of modelling the fluid-structure interaction of airway obstructions.

Chapters 5 and 6 contain selected results from the appended papers alongside more recent results that have not been published, with Chapter 5 focusing on the findings related to phonation and Chapter 6 on the observed effects due to fluid-structure interaction as present in the upper respiratory tract.

Finally, Chapter 7 concludes the work performed and states the most important findings. An outlook of possible future work is given.

The second part of the thesis includes the following papers:


Paper 4: L. Schickhofer, M. Mihaescu. *Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes.* To be submitted.


The numerical simulations, post-processing, and writing presented in this thesis have been carried out by the author, unless stated otherwise.

January 2019, Stockholm

*Lukas Schickhofer*
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Preface</td>
<td>viii</td>
</tr>
<tr>
<td><strong>Part I. Overview and Summary</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Chapter 1. Introduction</strong></td>
<td>3</td>
</tr>
<tr>
<td>1.1. Anatomy of the Human Upper Airways</td>
<td>4</td>
</tr>
<tr>
<td>1.2. Physiological Function</td>
<td>5</td>
</tr>
<tr>
<td><strong>Chapter 2. Background and Motivation</strong></td>
<td>11</td>
</tr>
<tr>
<td>2.1. Speech Production</td>
<td>11</td>
</tr>
<tr>
<td>2.2. Conditions Affecting Voice Generation</td>
<td>14</td>
</tr>
<tr>
<td>2.3. Snoring and Obstructive Sleep Apnea</td>
<td>16</td>
</tr>
<tr>
<td>2.4. Research Objectives</td>
<td>20</td>
</tr>
<tr>
<td><strong>Chapter 3. Modelling of Voice Production</strong></td>
<td>25</td>
</tr>
<tr>
<td>3.1. Reduced-Order Models</td>
<td>25</td>
</tr>
<tr>
<td>3.2. Advanced Approaches</td>
<td>26</td>
</tr>
<tr>
<td>3.3. Sound Sources and Acoustic Analogies</td>
<td>34</td>
</tr>
<tr>
<td>3.4. Rigid Glottal Model</td>
<td>37</td>
</tr>
<tr>
<td>3.5. MRI-Based Vocal Tract Model</td>
<td>39</td>
</tr>
<tr>
<td>3.6. Verification and Validation</td>
<td>42</td>
</tr>
<tr>
<td><strong>Chapter 4. Modelling of Airway Obstructions</strong></td>
<td>49</td>
</tr>
<tr>
<td>4.1. Fluid-Structure Interaction</td>
<td>49</td>
</tr>
<tr>
<td>4.2. Elastic Element in Channel Flow</td>
<td>50</td>
</tr>
<tr>
<td><strong>Chapter 5. Phonation and Glottal Aerodynamics</strong></td>
<td>53</td>
</tr>
<tr>
<td>5.1. Generation of Vowel Sounds</td>
<td>53</td>
</tr>
<tr>
<td>5.2. Glottal Flow and Supraglottal Flow Structures</td>
<td>55</td>
</tr>
</tbody>
</table>
Chapter 6. Aeroacoustics of a Simplified Soft Palate
  6.1. Flow Instabilities and Vortex Generation 65
  6.2. Vibrations and Dominant Frequencies 67

Chapter 7. Conclusions and Outlook 71
  Scientific Contributions 73
  Challenges and Limitations 73
  Future Work 74

Acknowledgements 76

References 78

Part II. Papers 87

Summary of the Papers and Author Contributions 89

Paper 1: Influence of changes of the glottal waveform on vowel production 95

Paper 2: On direct aeroacoustics calculations of the vocal tract 109

Paper 3: Compressible flow simulations of phonation using realistic vocal tract geometries 119

Paper 4: Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes 149

Paper 5: Aeroacoustics of an elastic element in unsteady flow of low Reynolds numbers 177
Part I

Overview and Summary
CHAPTER 1

Introduction

The human voice is the most perfect instrument of all.

Arvo Pärt

We are exposed on a daily basis to sounds of different origin. The generation of acoustic waves might be unwanted and disruptive such as in the case of noise generated through industrial processes, jet engines and airplanes, road vehicles, as well as air conditioning systems. However, the production of particular sounds might also be highly desired as in the case of musical instruments.

Also our own bodies are the source of sound - both wanted, such as in voice production for talking and singing (i.e. phonation), and unwanted, as a result of flow-induced vibrations and oscillations of soft tissues in the respiratory tract (e.g. snoring, wheezing). The centre of this sound generation are the human upper airways, which form a complex biological system, responsible for two important functions in our body: On the one hand, they contain the voice box necessary for phonation. Thus, they allow us to express ourselves by the articulation of sound and speech, which is an exclusive human trait. In addition to being the primary means of communication through verbal expression, our voice enables us to show emotions as well as to create music. On the other hand, the upper airways consist of the primary conduits for the transport of air to the lungs and are vitally important for the supply of oxygen.

Due to the interaction of rigid and elastic tissue with the in- and exhaled air, research efforts for a better understanding of the functioning of the upper airways are multi-disciplinary and encompass the fields of fluid mechanics, solid mechanics, and acoustics. However, research concerned with the functioning of the upper airways from a biomechanical perspective is a relatively young field. Since the 1940s engineers and physicists have been trying to uncover the physical processes of speech production and fluid-structure interaction in the upper airways.

This chapter gives a brief overview of the general anatomy and introduces the medical terms used in the field (cf. Sec. 1.1). This is important for the terminology used in the modelling approaches presented in Chapters 3 and 4 and for the description of selected results in Chapters 5 and 6 of this thesis. Additionally, the modes of operation for both phonation and respiration are
explained and the biological functions of the upper airways are described (cf. Sec. 1.2).

1.1. Anatomy of the Human Upper Airways

Physicians typically refer to one of the anatomical planes or axes, as shown in Fig. 1.1, when describing functions of the human body or pathological conditions. These are also used throughout this thesis.

![Anatomical Planes](image)

**Figure 1.1.** Sketch of the various anatomical planes, axes, and directions with respect to the human upper airways.

Fig. 1.2 gives an overview of the critical parts of the human upper airways, as viewed in the sagittal plane section. Starting from the nose and mouth opening, the airways consist of the nasal and oral cavities that are separated by the hard palate, which is a bony, rigid structure. Directly attached to the hard palate is the soft palate, an elastic, deformable appendage with its short tail end, the uvula. The first part of the upper airways is also referred to as pharynx, which is usually divided into nasopharynx (between nose and hard palate), oropharynx (from the hard palate to the epiglottis), and laryngopharynx (between the tongue and the larynx) (Ayappa & Rapoport 2003). The second part is called larynx and is more commonly known as the voice box. It embraces the essential parts for phonation such as the vocal folds, vestibular folds, and epiglottis. The larynx extends from the upper border of the epiglottis to the trachea. The vocal folds, which are a multi-layered organ, consist of three layers of tissue: the thyroarytenoid muscle, the lamina propria, and the epithelium (Hirano et al. 1981). The tension in the vocal folds is adjusted by the intrinsic laryngeal muscles - the thyroarytenoid, posterior cricoarytenoid, cricothyroid, and lateral cricoarytenoid. They are symmetric and enclose a
constriction, which is also referred to as glottis. The region just above the glottis is called supraglottis, while the region below it and towards the lungs is also called subglottis. The vestibular (or ventricular) folds, which are also called false vocal folds, form a second orifice close to the true vocal folds. They are located at a distance of approximately 3.7 mm to 7.5 mm from the vocal folds for men and 2.3 mm to 5.7 mm for women (Agarwal et al. 2003; Agarwal 2004). In the supraglottal domain lies the epilarynx tube, which contains the ventricular folds and the epiglottis. The upper airways are terminated by the opening of the trachea, which resembles a rigid tube and further guides the air towards or from the lungs.

1.2. Physiological Function

1.2.1. Larynx

In addition to the purpose of air passage to the trachea, the larynx is responsible for phonation. The vocal folds constitute the centre and most important part of the larynx. They work as a gateway between the subglottal region below the vocal folds with a particular lung pressure and the supraglottal region above with atmospheric pressure. Apart from being the central organ of phonation, the vocal folds also ensure that intrusion of foreign particles into
the trachea is prevented (Gehr 1994). Their main purpose, however, is the modulation of the pressure drop via the glottis in order to obtain a periodic source signal for voice generation. This is achieved by periodic opening and closing of the vocal folds that are excited to sustained oscillations. In order for phonation to be initiated, the vocal folds must be closed and the subglottal pressure must exceed the supraglottal pressure by a threshold value. Air being forced from the lungs is then pushing the vocal folds apart until the elastic forces in the tissue are bringing them back together again. This flow-induced vibration leads to successive adduction and abduction of the vocal folds. As the vocal folds oscillate, a mucosal wave is transmitted along the tissue surface, leading to an orifice geometry that varies in time from a convergent, to a parallel, and eventually to a divergent constriction before the cycle is repeated again, as depicted in Fig. 1.3. Thus, the flow through the glottis becomes intermittent during voiced speech, generating a pulsatile, turbulent glottal jet. The oscillation frequency of the vocal folds during phonation ranges from 80 to 220 Hz, with the male voice lying in the lower frequency range due to larger dimension of the larynx, and female voice and children’s voice occupying the higher frequency band. The glottal source signal is propagated along the vocal tract and modified by the shape of the airways and cavities as outlined in Sec. 1.2.3. However, periodic vibration is obtained only in certain conditions of the
transglottal pressure drop, the glottal opening, and muscle tension. In the case of unvoiced speech (e.g. whispering), the vocal folds remain static and sound is solely generated by the vortical structures propagated through the glottal jet and its shear layers.

As described in Sec. 1.1, the false vocal folds form a second constriction close to the true vocal folds. However, unlike the true vocal folds, they are static and not actively participating in phonation. Directly bordering to the vestibular folds and at the upper end of the larynx is the epiglottis, which has the important function of separating the larynx from the esophagus and prohibits food and saliva from entering the airway to the lungs during the process of swallowing. It also defends the lower respiratory tract against infection. The contribution of false vocal folds and epiglottis to phonation is still largely unclear and under debate. It has been shown in experiments based on excised larynges that the addition of false vocal folds and epiglottis increased the glottal flow resistance. Furthermore, Alipour et al. (2007) found the vestibular folds to be responsible for the low-frequency component in the acoustics as well as to increase the sound intensity at low frequencies of 50 to 100 Hz. Additionally, computational studies suggest an impact of the gap between false vocal folds on both the glottal jet and transglottal resistance (Mihaescu et al. 2013). In certain musical styles, like the Mongolian and Tibetan throat singing, as well as Sardinian chanting, the false vocal folds play a significant role and lead to an effect referred to as period doubling, where they oscillate with half the fundamental frequency of the true vocal folds (Fuks et al. 1998; Bailly et al. 2008, 2010).

1.2.2. Pharynx

In humans, the pharynx is part of both the conducting domain of the respiratory system as well as of the digestive system. While the nasal cavity is crucial
for cleaning and pre-conditioning of inspired air, the nasopharynx guides it to the throat. Both hard and soft palate are covered by respiratory mucosa tissue that further cleans, warms, and moistens the air. They have an average total surface area of 140 cm² to 160 cm² for an adult. The soft palate is composed of soft, elastic tissue that can easily collapse or be excited to strong vibration under pathological respiratory conditions (cf. Sec. 2.3). Hard and soft palate can be seen through the mouth opening, as depicted in Fig. 1.4. Oropharynx and laryngopharynx serve as a passage for both air and food. Therefore, the epiglottis temporarily covers the entrance to the larynx to suppress food from entering the respiratory tract to the trachea and lungs. The pharynx has a tubular shape and has an approximate total length of 12 to 14 cm (Gehr 1994). In large parts, the upper respiratory tract is a system of pipes transporting air to the lungs and the alveoli. However, unlike the trachea and the bronchii, which can be found between the larynx and the lungs, the upper airway is elastic, compliant, and collapsible.

The upper airway is, much like most vessels in the human body, subject to a certain transmural pressure, which is the difference between internal and external pressure. A collapsible tube model of comparable elastic properties shows a very similar behaviour to the airway tract during collapse (Heil & Jensen 2003). If the transmural pressure locally drops below zero, the tube starts to buckle non-axisymmetrically. In this state, it is highly sensitive to further changes of the internal pressure and can easily deform (Heil & Hazel 2011).
The self-excited oscillations of a collapsible tube have been investigated theoretically, experimentally, and numerically. Schwartz & Smith (2013) showed that the pharynx behaves like a Starling resistor, a model of an elastic, collapsible tube surrounded by a chamber at a critical pressure $P_{crit}$ that simulates tissue pressure, as depicted in Fig. 1.5. Studies on Starling resistors have identified several different types of flow-induced oscillations, including high-frequency flutter and low-frequency waves (Bertram 2003). Despite the considerable amount of studies conducted in this research area, the understanding of processes such as the generation and propagation of instabilities in single-phase flow through elastic tubes, leading to wall oscillations as they occur in the upper airway, is still incomplete (Grotberg & Jensen 2004; Heil & Hazel 2011).

1.2.3. Upper Cavities

In addition to its function as airways, the upper cavities, in particular the nasal and oral cavities, form the vocal tract together with the pharyngeal ducts and are thus also involved in voice generation. In this context, speech production can be regarded as a two-step process: The larynx, which is further explained in Sec. 1.2.1, acts as voice source emitting a signal of constant base frequency, while the vocal tract consisting of the airways above the vocal folds can be seen as a passive resonator or acoustic filter, which modifies the signal and amplifies it at its resonance frequencies. Therefore, different sounds and vowels can be generated by a particular vocal tract geometry with the fundamental frequency $f_0$ being constant. This happens through muscular flexion and extension allowing the dynamic change of the upper airway volume during speech, as well as the positioning of the tongue. The various cavities of the vocal tract lead to the occurrence of particular resonance modes in the speech spectrum.

Fig. 1.6 shows the internal upper airway geometries extracted from MRI scans of a healthy subject vocalising specific vowels. It can be seen that already slight changes in the resonance volume of the upper airway cavities at constant source frequency can cause diverse sounds being propagated into the acoustic far field. Speech is then produced by the continuous manipulation of the upper airway geometry. This process is powered by a supply of airflow from the lungs. A detailed description of the voice generation process and its open questions is given in Sec. 2.1.
Figure 1.6. Vocal tract geometries of Finnish vowels. The surface data has been provided via MRI measurements by the Speech Modelling Group at the Department of Mathematics and Systems Analysis, Aalto University.
CHAPTER 2

Background and Motivation

The main motivation for this work is the role of flow-induced sound generation and successive sound propagation during phonation under realistic conditions. The insights gained through numerical studies can further advance the understanding of the function of the upper airways under normal circumstances.

Another important reason is the large prevalence of vocal and respiratory disorders affecting the common population. A possible outcome is to supply medical specialists with alternative advice on therapeutic interventions besides the common clinical methods that are largely based on experiences of physicians.

In the following, background information is given on the basic mechanisms of sound production during phonation and on the state-of-the-art of the voice sciences (cf. Sec. 2.1). Additionally, the most wide-spread airway diseases and pathologies and their impact on the general population is described (cf. Sec. 2.2-2.3). Eventually, the research goals of this project are derived from these sections (cf. Sec. 2.4).

2.1. Speech Production

Speech is a crucial part of interpersonal communication. In this context, our voice is important for the expression of information, as well as emotional states. The human voice production is often modelled via a source-filter approximation, with the glottis acting as an acoustic source and the vocal tract as a filter (Titze & Alipour 2006; Story 2015). Through the regular opening and closing of the vocal folds, the volumetric flow rate is pulsed and results in a glottal volume velocity waveform, also referred to as glottogram. Several analytical models have been suggested to describe the behaviour of the glottal source. Early examples are, among others, the Fant model, the Liljencrantz-Fant model, the Fujisaki-Ljungqvist model, and the Rosenberg model (Fant 1979; Fant et al. 1985; Fujisaki & Ljungqvist 1986; Rosenberg 1971). Fujisaki & Ljungqvist (1986) give a comprehensive overview and analysis of various glottal waveform models. The shape of the glottogram has a large impact on the radiated sound. Gauflin & Sundberg (1980) could demonstrate that glottogram amplitude and peak amplitude of the differentiated glottogram are determining the sound pressure levels of vowel articulations with high accuracy.
Reflections and the resulting constructive and destructive interference of acoustic waves causes the periodic signal from the vocal folds to be modulated by the upper airway volume in such a way that only certain distinctive peaks remain in the envelope of the Fourier-transformed pressure fluctuations. These peaks are referred to as formants. Their position in the Fourier spectrum defines the articulated sound and in the case of vowels it is particularly the first two formants, \( F_1 \) and \( F_2 \), that identify the spoken tone. The spatial distribution and cavity association of individual formants have been first investigated by Fant & Pauli (1974) using one-dimensional models. Thus, it was possible to connect particular domains of the vocal tract with the occurrence of formants of certain frequencies. Subsequently, also three-dimensional wave equations have been applied to study this phenomenon (Stevens 2000).

*Sound-Generating Mechanisms.* Over time voice scientists identified the origins of the spoken sound and the underlying physical processes. Moreover, the production of our voice and thus speech in general can be broadly defined by three sound-generating mechanisms (Becker *et al.* 2009; Döllinger *et al.* 2016):

(i) Volume-modulated sound,
(ii) Vortex-induced sound,
(iii) Vibration-induced sound.

Thus, the overall sound signal can be divided into these three components. Each component has a varying contribution and their maxima of intensity lie in characteristic frequency ranges (Alipour *et al.* 2011). The volume-modulated sound is caused by the periodic opening and closing of the glottal constriction by the vocal folds (cf. Sec. 1.2.1). This leads to a pulsation of the airflow from the lungs and thereby of the subglottal pressure. As a result of the continuous oscillation of the vocal folds during voiced speech, most of the emitted acoustic energy is contained in the so-called *fundamental frequency* of phonation \( f_0 \), which is the frequency of the glottal pulse, and its upper harmonics. Hanna (2014) found the fundamental frequency to be proportional to the subglottal pressure through experiments on excised human larynges. The vortex-induced sound is produced through coherent structures being shed in the intraglottal region between the vocal folds, or in the supraglottal domain just downstream of the glottis (Mihaescu *et al.* 2010; de Luzan *et al.* 2015). These can be generated by shear-layer instabilities of the glottal jet or the sudden expansion of the vocal folds such as in the case of their divergent configuration during the closing phase (cf. Fig. 1.3). The glottal jet, together with the unsteady forces acting on the vocal folds, is directly responsible for the sound generation in the epilarynx (Mittal *et al.* 2011). Moreover, it is believed that the developed flow structures during the closing phase of the phonation cycle have an impact on voice quality (Khosla *et al.* 2009; Mihaescu *et al.* 2010). Finally, the vibration-induced sound is due to the structural motion of the vocal fold tissue leading to the displacement of air around it and thus causing the propagation
of sound waves away from it. However, this component is considered to be of much smaller amplitude than the other two and is typically neglected in most modelling approaches.

**Harmonic and Inharmonic Component.** The dominant mechanism of sound production at the fundamental frequency as well as at its upper harmonics is the volume modulation (Klatt & Klatt 1990). It is the defining source of sound at lower frequencies. At higher frequencies, however, the broadband noise component caused by vortices in the supraglottal region becomes more and more important. The emitted acoustic frequencies are indirectly proportional to the characteristic size of the vortices. Smaller vortical structures are thus producing sound at higher frequencies. Eventually, this mechanism dominates the sound sources at high frequencies. The acoustic voice spectrum as it is perceived by a listener can therefore be roughly separated into a harmonic component and an inharmonic one. The harmonic component consists of the fundamental frequency $f_0$ of vocal fold motion as described in Sec. 1.2.1 and its upper harmonics ($2f_0, 3f_0, 4f_0,...$). It is easily identifiable in the voice spectrum through its regularly distributed peaks in the low-frequency range up to about 3 kHz, which marks the approximate position of the third formant $F_3$ in the spectrum, as shown for the example of vowel [a] in Fig. 2.1. Furthermore, it is the harmonic part of the spectrum that carries the most important acoustic information for the identification of speech sounds. This information is given by the first formants as peaks of the envelope. At higher frequencies above several kilohertz, however, the irregular, inharmonic broadband sound component through supraglottal flow structures gains in amplitude relative to the upper harmonics of the source frequency. Zhang et al. (2002b) and Howe & McGowan (2007) could show through application of the quasi-steady approximation of aeroacoustics that the contribution of vortices and turbulence in the laryngeal airflow to the resulting sound field becomes important at frequencies above 2 kHz. Apart from the acoustic impact of supraglottal flow structures, they also significantly influence the orientation and deflection of the glottal jet and large-scale recirculation zones downstream of the glottis (Kniesburges et al. 2013).

**Voice Quality and Characteristics.** As stated by Sundberg (1994), there are two important characteristics of voice sounds: *Vowel quality* on the one hand, which is mainly influenced by the position of the first two formants in the frequency spectrum; and *voice quality* on the other hand, which depends mainly on the high-frequency information of the spectrum. This individual voice quality might be impacted by the position of higher formants, but also broadband sound through turbulent fluctuations that could mask these tonalities. Kreiman & Gerratt (2012) further emphasized the importance of the inharmonic broadband component to voice quality and the perceptual importance of noise. The relative broadband sound in terms of the noise-to-harmonic ratio (NHR) was
Figure 2.1. Voice spectrum of the spoken vowel [a] in the near and far field. The approximate ranges of the harmonic and inharmonic component are indicated. The positioning of point probes $P_1$ to $P_4$ is shown in Fig. 5.9.

found to be a strong indicator of breathiness of a voice (de Krom 1995; Hillenbrand 1988).

Although there is significant academic interest in closing the gaps in understanding of the voice generation process, there is also an increasing interest in the clinical investigation of pathological conditions of the phonatory system, as further described in Sec. 2.2. Realistic, patient-specific geometries that can be obtained through magnetic resonance imaging (MRI) data are required to allow for reliable conclusions regarding the effects of disease-related changes concerning the vocal tract.

2.2. Conditions Affecting Voice Generation

The ability to speak and express oneself by phonation is a basic human feature. However, the capacity to do so can be considerably compromised by diseases affecting the upper airways. The majority of people is able to perform regular voice production as described in detail in Sec. 2.1. However, up to 30% of the human population are estimated to develop a voice production disorder at some point in their life (Mittal et al. 2013). Roy et al. (2004) and Sliwinska-Kowalska et al. (2006) found that certain professions with extensive voice use, like teachers and professors, show an occurrence of voice disorders of as high as 60%. At any given time, approximately 7% of the general population are suffering from some kind of vocal disorder. Although most of these conditions are short-lived, with less than four weeks in duration, many lead to severe permanent
effects on job performance, occupational voice use, and social engagement (Roy et al. 2005). Women have a slightly higher incidence of voice disorders. It has been hypothesized that the reason for this lies in the differences in laryngeal anatomy: Women have shorter vocal folds vibrating at a higher fundamental frequency. As a result, there is less tissue mass available to dampen higher vibratory force, which is released at the impact during adduction. Conditions of the vocal folds and upper airways can directly influence the fundamental frequency of the periodic movement of the glottis as well as the waveform of the source signal. This could further impair a patient's ability to excite resonances of the vocal tract and generate vowels. Laryngeal pathologies can significantly contribute to the occurrence of non-linear behaviour during voice production.

The most prevalent of voice disorders is unilateral (partial) or bilateral (total) vocal fold paralysis or paresis (VFP). Patients of VFP suffer from tissue weakness leading to insufficient closure of the vocal folds. If only one side of the glottis is affected, this causes an asymmetry in the vibration pattern, which relates to hoarseness, increases voice fatigue, and decreased loudness. Laryngeal paralysis has been linked to a considerably larger opening quotient of affected patients when compared to normal subjects (Hanson et al. 1988). The opening quotient gives the ratio of the time when the glottis is open to the total pulsation period. Additionally, the intraglottal flow field happens to be asymmetric, which leads to the glottal jet attaching predominantly to the immobile side of the constriction affected by paralysis (Erath & Plesniak 2010). Also biphonation might occur, in which case the two vocal folds are vibrating at two different frequencies. Xue et al. (2010) showed that the asynchronous movement pattern caused by unilateral vocal fold paralysis leads to a higher subglottal pressure necessary for phonation onset and that the resulting fundamental frequency is given by entrainment of the two eigenmodes of the vocal folds. Overall, the asymmetric behaviour of the tissue and the flow as a result of vocal fold paralysis causes non-linear effects and in some cases even chaotic frequency spectra.

Further examples of diseases affecting vocal fold closure and functioning are polyps, papilloma, cancer at the glottis, or physical trauma at the voice box (Mathieson 2013). All these conditions are due to lesions at the vocal fold surfaces, which in turn inhibit complete closure. Consequently, a higher subglottal pressure is required for a sustained vibration of the vocal folds and the occurring lesions introduce additional vortex shedding at the glottis. The occurring vortices disrupt the glottal aerodynamics and have been shown to alter the location of flow separation downstream (Erath & Plesniak 2012). Tab. 2.1 shows the physical causes behind the most common symptoms affecting the ability to speak.

The most wide-spread surgical intervention in cases of incomplete glottal closure is the laryngoplasty, in particular the medialisation thyroplasty. In this procedure, the immobile vocal fold is adjusted by a silicon or plastic implant.
Table 2.1. Symptoms of common vocal disorders and their biomechanical cause. Data based on the work by Mittal et al. (2013).

<table>
<thead>
<tr>
<th>Vocal condition</th>
<th>Biomechanical origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoarseness</td>
<td>Aperiodic vocal fold vibration</td>
</tr>
<tr>
<td>Vocal fatigue</td>
<td>High subglottal pressure</td>
</tr>
<tr>
<td>Breathy voice</td>
<td>Incomplete glottal closure</td>
</tr>
<tr>
<td>Vocal fold nodules</td>
<td>Excessive contact stress on the vocal folds</td>
</tr>
<tr>
<td>Biphonation</td>
<td>Two independent vibrational frequencies</td>
</tr>
</tbody>
</table>

such that contact between the vocal folds and therefore the closing phase of the glottal cycle is reestablished. The positive outcome of this surgery is highly dependent on the correct placement of the implant, for which there are no exact, objective criteria. Also the vestibular folds have been found to develop pathological behaviour, called ventricular dysphonia (Nasri et al. 1996). In this case the false vocal folds are excited to oscillation during phonation and disturb the regular movement of the true vocal folds due to back-coupling.

In speech biomechanics, there is considerable research effort dedicated towards finding ways to treat patients with disorders in the upper airways affecting their ability to talk. However, the possibilities to examine the effects of conditions in these sensitive areas are limited due to multiple reasons. First, the vocal tract is a confined and complex cavity that makes in-vivo measurements of its internal dimensions and elastic properties a difficult or impossible task. Additionally, phonation implies a complex unsteady flow-structure-acoustics interaction. It is a highly dynamic process, during which the behaviour of the upper airways is hard to monitor. Thus, when dealing with diseases of the upper airways and the vocal tract, physicians are often limited in their options to assess the effect a particular surgical procedure might have on the patient. There is no method to estimate a priori what the outcome of a therapeutic intervention could be for the patient’s ability to speak. This poses, together with the large prevalence of laryngeal pathologies in the common population, a major motivation for the study of flow and acoustics in the vocal tract. As a result, such research will progress the understanding of cause-effect relationships of phonation and help to gain insight into the fundamental mechanisms generating the wide range and richness of the human voice.

2.3. Snoring and Obstructive Sleep Apnea
Breathing is a basic vital function, which can be significantly compromised for patients with airway disorders. The upper respiratory tract resembles a system
of deformable tubes of low elastic modulus. Therefore, its opening is dependent on the air pressure and pressure inside the tissue, as discussed in Sec. 1.2.2. Due to muscle relaxation during sleep, this can lead to collapse of the airways, which poses a severe danger for affected patients. Hypopnea means partial closing of the pharyngeal airway, while apnea implies total closing. Obstructions of the upper airways during sleep are a common occurrence affecting large parts of the population of developed countries. In particular obstructive sleep apnea (OSA) and related conditions affect both life expectancy and life quality of patients considerably. The prevalence of OSA depends on the applied definition. Typically, an average number of 10 apneic and hypopneic episodes per hour of sleep is considered as strong indication of OSA (Silverberg et al. 2002). The number of such events per sleep hour is generally referred to as apnea-hypopnea index (AHI), or respiratory disturbance index. If the common definition of an AHI of 10 or more is applied for the classification of sleep apnea, it can be shown that about 10% of people in the age group of 30-60 years have OSA, with 15% of men and 5% of women (Young et al. 1993), making it the most widespread obstructive airway disorder with more than 12 million affected people in the United States alone. Although sleep apnea can manifest at any age, the incidence of cases rises both with age and gain in bodyweight (Fietze et al. 2011).

Alongside the more obvious complications that can arise due to airway constrictions during sleep, like choking or inhibition of air transport to the lungs leading to low brain oxygen saturation, OSA has been linked to essential hypertension, heart failure, and premature death (Lavie et al. 2001; Silverberg et al. 2002; Naughton 2006; Oldenburg et al. 2007).

Besides being a harmful condition for the affected patients themselves, OSA has been identified as a high risk factor for motor vehicle accidents and thus represents a potential threat for the general public (Tregear et al. 2009; Philip et al. 2010). This is due to the excessive daytime sleepiness, drowsiness, and lack of concentration resulting from unrestful sleep. In recent years, meta-analyses have been carried out that evaluated the relative risk factor of accident involvement of drivers with certain medical conditions: While mental disorders lead to a relative risk of 1.72, drugs and medicine to 1.58, and alcoholism to 2.00, sleep apnea and narcolepsy has a risk factor of 3.71, with most of it due to sleep apnea (Mäkinen et al. 2003). This means that a patient of OSA is almost four times as likely to be involved in a motor vehicle accident as an ordinary person.

A screening for OSA before both receiving and renewing one's driving license is under constant discussion by a specialised working group of the European Union (McNicholas 2013). Apart from the considerable human toll and harm to health caused by OSA and related sleep disorders, there is also a significant financial burden attached to it. In the United States, the indirect costs of sleep disorders to the general public have been estimated by the American National Commission on Sleep
Disorders Research to be as high as 15 billion US dollar per year (Shone et al. 1998). In particular untreated sleep apnea is found to accumulate about 3.4 billion US dollar of additional medical costs. Additionally, OSA is estimated to be twice as expensive for the health care system prior to diagnosis as afterwards, stressing the importance of quick and reliable diagnostic methods (Kapur et al. 1999). Since patients of OSA typically show no detectable respiratory abnormality while awake, the clinical diagnostic method of choice is a so-called polysomnography, which involves monitoring vital function and respiration frequency among other parameters of the patient while sleeping. However, sleep monitoring is not only a time-consuming process for both patient and physician, it is also expensive with an average health care cost of 390 Euro per study and person in the European Union. The range of costs in different countries is from 700 Euro in Finland and Germany to 180 Euro in Greece, with Sweden placed around the mean (Escourrou et al. 2000). The main diagnosis of performed polysomnography in Europe is sleep apnea in adults, confirming its standing as the most prevalent sleep disorder. Obstructional diseases of

<table>
<thead>
<tr>
<th></th>
<th>Soft palate length (cm)</th>
<th>Soft palate area (cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal subjects</td>
<td>4.1 (±0.4)</td>
<td>3.7 (±0.5)</td>
</tr>
<tr>
<td>Simple snorers</td>
<td>4.5 (±1.1)</td>
<td>4.8 (±1.2)</td>
</tr>
<tr>
<td>OSA patients</td>
<td>4.7 (±0.7)</td>
<td>5.6 (±1.3)</td>
</tr>
</tbody>
</table>

the upper airway usually lead to audible sound being generated by tissue oscillations in the acoustic near field. Elastic structures subjected to unsteady flow of low Reynolds number tend to exhibit strong oscillatory behaviour under certain conditions of the pressure fluctuations. This occurs, if the frequency of the shed vorticity in the flow lies in the vicinity of the natural frequency of the structure. The amplitude of the vibrational response then depends heavily on the elastic modulus. In extreme cases resonance can be obtained and the magnitude of the displacements reaches a maximum. The resulting structural oscillations emit acoustic waves of significant sound power. Vibrations at the domain wall can cause noise generation of a broadband, noise-like character in the high-frequency range (> 500 Hz), which is typically referred to as pharyngeal or non-palatal snoring. However, the dominant sound source of snoring is the oscillating soft palate, an elongated structure of small Young’s modulus separating the oral from the nasal cavity, as described in Sec. 1.1, with the pitch of the snoring sound in the low spectral bands (< 500 Hz) (Huang et al.
This process is called palatal snoring. The soft palate tends to have a higher length and area for obstruction cases of increasing severity, as demonstrated in Tab. 2.2 for sample groups of normal subjects, pathological snorers, and OSA patients. The longer the soft palate, the higher the probability of partial or total occlusion of the respiratory channels.

Snoring source identification indicated that the origin of noise-generating vibrations are mainly the tongue, the soft palate, and the nasal cavity, with the highest sound power being emitted at the resonance frequencies of the soft palate (Liu et al. 2007). When both nose and mouth are open and an inflow occurs over upper and lower surfaces of the soft palate, an intermittent obstruction of the airways in the nasal and oral cavities takes place. If the mouth is closed, the soft palate temporarily obstructs the nasopharyngeal channel (Bertram 2008). In some cases snoring sound might result from oscillations of the elastic pharyngeal airway itself. All these scenarios however include the excited vibration of flexible parts of the upper airways leading to a considerable noise generation. The measurement and characterisation of tissue properties is widely considered a crucial obstacle towards the accurate simulation of patient-specific cases. Constitutive laws and equations of state for the modelling of the tissue behaviour have been based on experimentally determined parameters, which show large discrepancies depending on source and applied methodology, as demonstrated by data from different publications in Tab. 2.3. Additionally,

<table>
<thead>
<tr>
<th>Model</th>
<th>Young's modulus $E$ (Pa)</th>
<th>Poisson's ratio $\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birch &amp; Srodon (2009)</td>
<td>585-1409</td>
<td>0.22-0.54</td>
</tr>
<tr>
<td>Cheng et al. (2011)</td>
<td>7539</td>
<td>0.49</td>
</tr>
<tr>
<td>Zhu et al. (2012)</td>
<td>3000</td>
<td>0.49</td>
</tr>
<tr>
<td>Wang et al. (2012)</td>
<td>25000</td>
<td>0.42</td>
</tr>
</tbody>
</table>

the tissue surrounding the upper airways shows different stiffness and resistance depending on its active or passive state. The muscle activity and elasticity during sleep stops, causing relaxation of the tissue and making it more likely to collapse or be excited to vibrations.

It has been found in population studies that up to 81% of middle-aged men snore for more than 10% of the time between bedtime and getting up (Bearpark et al. 1995). Conservative estimates of snoring as a prevalent disorder suggest that it affects 20-40% of the general population. Although there is a strong correlation between obstructive sleep disorders and snoring, not all snorers are
automatically OSA patients. An estimated 10% of snorers are considered to be at risk to OSA (Bertram 2008; Chouly et al. 2008; Rasani et al. 2011). The role of the nasal cavity in snoring and sleep-disordered breathing, as well as its coupling to other parts of the upper airway, has yet to be understood (Rappai et al. 2003; Pevernage et al. 2005). The change in nasal airflow due to obstructions during sleep and the accompanying rise in airway resistance might trigger OSA in particular cases. However, the causality between nasal constrictions and the development of sleep-disordered breathing is not clear.

Different approaches exist for the treatment of OSA and heavy snoring. Usually, non-invasive therapeutic measures are preferred. The most common is the use of a continuous positive airway pressure (CPAP) device. Thereby, the patient wears a facial mask during sleep, which applies a pressure at the nose to keep the pharynx from collapsing. If non-invasive treatment plans fail, surgical procedures are considered. These can range from minimally invasive, like the pillar procedure, where small pillars are inserted into the soft palate to artificially stiffen the tissue, to highly invasive, like the uvulopalatopharyngoplasty, a downsizing or removal of the flexible parts around the uvula. Sher et al. (1996) conducted meta-studies based on long-term observation of patients and found that surgical procedures for the treatment of OSA are effective in less than 50% of cases. A high failure rate with surgeries turning out to be not helpful, or even detrimental for the patient are thus the reality. So far there is no possibility to assess the outcome of a surgical procedure at the airways a priori. However, by analyzing both flow and acoustics of the problem, valid assumptions can be made on how a particular airway configuration may trigger breathing problems or the generation of noise. Thus, it could be possible to anticipate the outcome of certain medical measures by means of numerical simulations in advance to actually performing them. A practical example is the study of efficacy of the mandibular advancement splint (MAS), a device worn during sleep that protrudes the lower jaw and thus potentially inhibits snoring and OSA events from happening (Zhao et al. 2013a,b).

2.4. Research Objectives

The main goal of this work is a better understanding of the physical processes in the upper airways related to the intended generation and propagation of sound, as it happens during speech (cf. Sec. 2.1). Furthermore, unintended sound generation due to airway obstructions and the coupling between the flow and flexible structures is explored (cf. Sec. 2.2-2.3). Particular emphasis is placed on the modelling of the voice generation process from the perspective of computational fluid dynamics. In the following, a brief outline of the main goals of the research project is given:
Sound Through Voice Production

- Test of reduced-order and more advanced models for acoustic simulations of the human upper airways during phonation. Initially, models of different complexity are assessed in terms of their ability to accurately predict acoustic frequencies and resonances of airway volumes. Up until now, no computational model exists that allows for a full simulation of the dynamic behaviour of the upper airway during the complex speech production process. This would require two-way coupled fluid-structure interaction allowing contact between the vocal folds in addition to an accurate assessment of flow, acoustics, and structural displacement.

- Verification and validation of the proposed numerical method for various vocal tract models.
  The numerical method chosen for this work is to be validated by the use of glottal and vocal tract geometries used in literature and experiments in order to compare the results with available data, such as profiles of mean surface pressure\(^1\) and mean velocity, as well as acoustic formant data\(^2\).
  The solution of the unsteady numerical flow simulations is compared for different cell counts and computational grid sizes for the investigated airway geometries. The discretisation errors shall be computed and minimised.

- Computation of the unsteady, intermittent flow through simulations of a realistic, subject-specific model of the vocal tract.
  Eventually, a computational tool shall be developed to allow for reliable prediction of flow and acoustics under realistic conditions. Historically, simplified tubular models based on cross-sectional area functions of the vocal tract under phonation have been studied in conjunction with wave reflection analogues (Stevens 2000; Titze & Alipour 2006). These simplifications have only little impact on the acoustic response for frequencies below 4–5 kHz. This is due to the fact that the sound waves in this low-frequency range propagate in a planar manner. However, as shown by Blandin \textit{et al.} (2015) and Arnella \textit{et al.} (2016), the higher-order modes of frequencies above that threshold are strongly dependent on the accurate geometrical representation of the vocal tract. Therefore, a full-scale airway model is constructed in this work, taking into account MRI images and

\(^1\text{cf. Scherer et al. (2001)}\)
\(^2\text{cf. Aalto et al. (2012)}\)
resulting surface files extracted by the Speech Modelling Group at the Department of Mathematics and Systems Analysis, Aalto University (Aalto et al. 2014).

- Investigation of the impact of flow instabilities and vortical structures in the intraglottal and supraglottal region on voice production. As stated by Mittal et al. (2013) and Zhang (2016), the relative importance of flow structures produced in the laryngeal and epilaryngeal domain is still unclear. Although Zhang & Neubauer (2010) argued that their influence on low-frequency voice production and vocal fold motion is negligible, it is commonly accepted that they significantly contribute to higher frequencies and modes of the voice spectrum (cf. Sec. 2.1). A goal of this work is the identification of dominant frequencies produced by unsteady laryngeal flow and its impact on far-field acoustics.

**Sound Through Airway Obstructions**

- Implementation of a method for strongly coupled fluid-structure interaction simulation of a simplified model of the pharyngeal passage. Simulations based on fluid-structure interaction with strong coupling are essential for the prediction of noise generation of flexible bodies under the influence of shear flow. Especially flow-induced body vibrations, which contribute to a significant part of the total sound output, need to be predicted accordingly. A model, which takes into account the two-way fluid-structure interaction between the airflow and the compliant parts of the respiratory tract is needed to assess the conditions of blockage or sound generation due to vibrations.

- Determine the sensitivity of flow instabilities, structural oscillations, and acoustics towards changes of critical parameters (e.g. Reynolds number $Re$, elasticity $E$).

  The different configurations are tested for varying inflow velocities and Reynolds numbers, taking into account various material parameters of patient-specific scenarios. The necessary soft tissue properties and constitutive models in literature show significant variations between studies (Birch & Srodon 2009; Cheng et al. 2011). Thus, it is necessary to study the impact of variations in quantities like the Young's modulus $E$ on the behaviour of the tissue.

- Identification and quantification of the acoustic sources inside simplified geometries of the upper airways by applying direct compressible simulations and acoustic analogies.
By studying the fluid-structure interaction and the acoustic response of the vibrating structures of the human upper airways to flow fluctuations, a correlation between dominant frequencies in the spectrum and the anatomical characteristics of the respiratory tract can be achieved. Resonance effects of flexible elements and the onset of aerodynamic sound generation need to be investigated in order to understand effects due to fluid-structure-acoustics interaction in the upper airways.

The findings of this work relevant to the above research objectives are presented as conclusions in Chapter 7. An overview of selected results is given in Chapters 5 and 6.
CHAPTER 3

Modelling of Voice Production

Reduced-order modelling is often applied for simulating the biomechanics of the upper airways. This is necessary due to their inherent complexity and the interaction of fluid, structure, and acoustics. It is particularly the case for speech acoustics, where the main interest lies in the prediction of resonance frequencies and sound pressure levels (cf. Sec. 3.1). A downside of this approach is the strong simplification and restriction to only planar wave propagation, while flow physics, instabilities, and turbulence are neglected.

For an enhanced physical understanding of the aerodynamically generated sound in the upper airways, more detailed approaches relying on resolving the fluid flow dynamics are required (cf. Sec. 3.2). Eventually, this chapter presents the aeroacoustic formulations and computational geometries of the vocal tract applied in this work (cf. Sec. 3.3–3.5), along with results of validation and verification of the methods (cf. Sec. 3.6).

3.1. Reduced-Order Models

When looking at the upper airway as a resonance body, it is possible to extract the dominant frequencies in the low-frequency range of the acoustic spectrum by solving a simplified problem based on the one-dimensional wave equation, describing the propagation of planar sound waves (Stevens 2000; Titze & Alipour 2006).

3.1.1. One-Dimensional Wave Equation

The one-dimensional homogeneous wave equation reads as

\[
\frac{1}{c_0^2} \frac{\partial^2 p_a}{\partial t^2} - \nabla_z^2 p_a = 0,
\]

for the acoustic pressure \( p_a \). The speed of sound at room temperature is \( c_0 = 343.21 \text{ m/s} \). A commonly known general solution of this equation is of d’Alembert type

\[
p_a (z, t) = p_a^+ (z - c_0 t) + p_a^- (z + c_0 t),
\]

which is a superposition of waves travelling in the positive and negative direction, indicated by the superscripts + and −, that are reflected by the solid
domain boundaries. The one-dimensional wave equation can be applied to geometries of different shapes in order to give the resonance frequencies of the particular airway volume. It is suited for simple symmetric geometries, where boundary conditions such as reflection and transmission coefficients can be easily specified. One such example is the wave reflection analogue for upper airways.

3.1.2. Wave Reflection Analogue

A common approach to model vocal tract acoustics is based on using the approximation of an axially symmetric tube for the upper airways (Liljencrants 1985a,b). Story et al. (1996) showed that the segments of the tube can be defined by vocal tract area functions that are extracted from an upper airway geometry. In this case the solution to the one-dimensional wave equation as introduced with Eq. (3.1) is a sum of d’Alembert waves propagating back and forth in the channel. Thereby, the resulting solution for the fluctuating pressure \( p_a \) is evaluated at each section interface of the tube. Space and time are discretised as \( k = z/\Delta z \) and \( n = t/\Delta t \), respectively. The acoustic resistance of the vocal tract is computed by taking into account the cross-sectional area \( A_k \) of its segments \( k \) as follows:

\[
R_k = \frac{A_k - A_{k+1}}{A_k + A_{k+1}}. \tag{3.3}
\]

Eventually, the solution for the acoustic pressure \( p_a \rightarrow p \) can be achieved as the recursive relationships

\[
P_{k,n}^- = (1 - R_k) P_{k+1,n-1}^- + R_k P_{k,n-1}^+,
\]

\[
P_{k+1,n}^+ = (1 + R_k) P_{k,n-1}^- - R_k P_{k+1,n-1}^-, \tag{3.4}
\]

in terms of incident and departing waves.

3.2. Advanced Approaches

Despite the good performance of reduced-order models for the prediction of resonance modes in the low-frequency domain of an airway geometry and their obvious advantage in computational costs, eventually only high-order modelling gives the necessary information about unsteady physical effects occurring during the process of phonation and acoustic wave interaction. This is especially the case for the calculation of high-frequency modes present in the speech signal that are due to vortex-induced sound. Alipour et al. (2011) and more recently Sadeghi et al. (2018) compared common simulation methods of human phonation and their numerical costs. An accurate resolution of the flow domain was emphasised alongside the importance of a realistic representation of the sub- and supraglottal regions, despite the considerable increase in required computational resources.
3.2.1. Three-Dimensional Helmholtz Eigenvalue Problem

The resonance frequencies of a full-scale, three-dimensional upper airway model, as it can be achieved through magnetic resonance imaging (MRI), can be obtained through solution of the homogeneous wave equation in three dimensions. This accounts not just for the predominantly planar wave propagation in the frequency range below 4–5 kHz, but also for other higher-order propagation modes (Blandin et al. 2015). The problem can be treated by rearranging the wave equation as a Helmholtz eigenvalue problem. Here, the wave equation is applied to the velocity potential $\phi = \phi (\vec{r}, t)$:

$$\left( \frac{\partial^2}{\partial x_i^2} - \frac{1}{c_0^2} \frac{\partial^2}{\partial t^2} \right) \phi = 0. \quad (3.5)$$

with the speed of sound in air $c_0$. By assuming the scalar potential to be separable,

$$\phi (\vec{r}, t) = \Phi (\vec{r}) f (t) = \Phi (\vec{r}) e^{ikc_0t}, \quad (3.6)$$

the wave equation (3.5) can be rewritten as an eigenvalue problem

$$\left( \frac{\partial^2}{\partial x_i^2} + k^2 \right) \Phi = 0, \quad (3.7)$$

with the derivative in the normal direction at the solid walls given by the Neumann boundary condition

$$\frac{\partial \Phi}{\partial n} = 0. \quad (3.8)$$

Eq. (3.7) is also referred to as Helmholtz equation. Its eigenvalues are a set of wave numbers $k_n$ for a particular domain, that are related to the Helmholtz frequencies by $f_n = \omega_n / (2\pi) = k_n c / (2\pi)$. These are the acoustic frequencies, at which a considered volume is resonating. By solving Eq. (3.7) for a particular domain, the dominant, energy containing peaks in the frequency spectrum are computed.

3.2.2. Numerical Flow Simulation

Computational fluid dynamics (CFD) is a fairly recent approach to voice acoustics that allows to compute acoustic pressures as a by-product of the numerical solution of the compressible Navier-Stokes equations. CFD has been applied within the field of biomechanics of the human upper airways with the aim to model and correctly predict the flow characteristics (Löhner et al. 2003).

For a direct computation of both the velocity field in the vocal tract and the acoustic pressure fluctuations in the near field, compressible flow simulations using the finite volume method are performed in this work. Therefore, the behaviour of the fluid is described by the Navier-Stokes equations in compressible form. The continuity equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_i)}{\partial x_i} = 0, \quad (3.9)$$
ensures conservation of mass in the fluid, while the momentum equation
\[ \frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j) = \frac{\partial}{\partial x_j}(-p \delta_{ij} + \sigma_{ij}), \] (3.10)
guarantees conservation of momentum. Energy conservation is ensured by
\[ \frac{\partial}{\partial t}(\rho e_0) + \frac{\partial}{\partial x_i}(\rho u_i e_0) = -\frac{\partial}{\partial x_i}(\rho u_i) - \frac{\partial q_i}{\partial x_i} + \frac{\partial}{\partial x_i}(u_j \sigma_{ij}). \] (3.11)
The position in Cartesian coordinates $x_i$, the fluid velocity $u_i$, and the heat flux $q_i$ are vector quantities with indices $i$, $j$, and $k$, while the time $t$, density $\rho$, dynamic viscosity $\mu$, and static pressure $p$ are the occurring scalar quantities. The total energy per unit volume $e_0$ is defined as the sum of the specific internal energy and the specific kinetic energy
\[ e_0 = e + \frac{1}{2} u_i u_i, \] (3.12)
with the internal energy being related to the enthalpy $h = c_p T$ via $e = h - p/\rho$, with the specific heat capacity of air at constant pressure $c_p$. If a Newtonian fluid is considered, the viscous stress tensor reads as
\[ \sigma_{ij} = 2\mu \left(S_{ij} - \frac{1}{3} S_{kk} \delta_{ij}\right), \] (3.13)
relating to the strain rate tensor $S_{ij}$ via the dynamic viscosity $\mu$.

Body forces like gravity are not considered in this study as their effect on airflow in the upper airways is negligible. In order to close the system of Eq. (3.9)–(3.11) for the four unknowns of the velocity components $u_i$ and the pressure $p$, an equation of state of the fluid is necessary: The airflow is modelled in the numerical simulations by the ideal gas law
\[ p = \rho RT, \] (3.14)
relating gas pressure $p$, with density $\rho$ and temperature $T$ via the specific gas constant of air $R = 287.06 \text{ J/(kgK)}$. In this work, the air is treated as an isothermal gas at a constant ambient temperature of $T = 300 \text{ K}$.

**Finite Volume Method.** Discretisation of the equation system (3.9)–(3.11) describing the fluid motions is performed by using the finite volume method. Thus, the differential quotients occurring in the above equations are transformed into difference quotients, leading to an algebraic system. All of the considered conservation equations are written in the form of a transport equation
\[ \frac{d}{dt} \int_{\Omega} \rho \phi d\Omega = \int_{\Omega} \rho \bar{u} \phi \cdot d\bar{A} + \int_{\partial \Omega} \Gamma \nabla \phi \cdot d\bar{A} + \int_{\Omega} S_{ij} \phi d\Omega, \] (3.15)
which is formulated for a control volume \( V \) and by applying Gauß’ divergence theorem

\[
\int_V \left( \nabla \cdot \mathbf{F} \right) dV = \int_{\partial V} \left( \mathbf{F} \cdot \mathbf{n} \right) dA, \tag{3.16}
\]

connecting the volume integral of the divergence of a vector field \( \mathbf{F} \) to the flux over the volume’s boundaries \( \partial V \).

Following the specifications of the used solver (Simcenter STAR-CCM+ User Guide 2018), the surface integrals are evaluated as the product of the value at the computational cell face centre and the cell face area:

\[
\int_A \mathbf{F}^{\phi} \cdot d\mathbf{A} \approx \sum_f \mathbf{F}^{\phi}_f \cdot \mathbf{A}_f. \tag{3.17}
\]

Furthermore, the volume integration is approximated by the second-order accurate product of the mean value of the source term at the cell centre and the cell volume:

\[
\int_V S^{\phi} dV \approx S^{\phi}_0 V_0 \tag{3.18}
\]

Eq. (3.15) gives the total transport of a scalar quantity \( \phi \) in four terms, which are solved in the following way:

- The left-hand side of Eq. (3.15) is the transient term, representing the change of the variable \( \phi \) in time inside the control volume \( V \). The term is discretised into time steps, where the current solution requires the solutions from previous time steps. A second-order temporal discretisation of the term is applied as

\[
\frac{d}{dt} \left( \rho \phi V \right) \approx \frac{3 (\rho \phi V)_{n+1} - 4 (\rho \phi V)_n + (\rho \phi V)_{n-1}}{2\Delta t}, \tag{3.19}
\]

using the solution of the previous two time levels.

- The convection term expresses the net rate of decrease of the property \( \phi \) due to convection and can be discretized as

\[
\left( \rho \mathbf{u} \phi \cdot \mathbf{A} \right)_f = \left( \mathbf{n} \phi \right)_f \approx \bar{m}_f \phi_f, \tag{3.20}
\]

with the mass flow rate \( \bar{m}_f \) at the face and the value of the fluid variable \( \phi_f \) at the face. The convective flux is then computed using a bounded central-differencing scheme

\[
\left( \bar{m} \phi \right)_f = \begin{cases} 
\bar{m}_f \phi_{FOU} & \text{if } \xi < 0 \text{ or } \xi > 1, \\
\bar{m}_f \left[ \sigma \phi_{CD} + (1 - \sigma) \phi_{SOU} \right] & \text{if } 0 \leq \xi \leq 1,
\end{cases} \tag{3.21}
\]

which uses a cell-face centre value based on different schemes, depending on the value of the normalised variable

\[
\xi \left( \mathbf{r}, t \right) = \frac{\alpha \left( \mathbf{r}, t \right) - \alpha_U}{\alpha_D - \alpha_U}, \tag{3.22}
\]

23
as a function of the nodal values $\alpha_D$ and $\alpha_U$ for downwind and upwind positions of the variable near a cell face $\alpha (r, t)$. The first-order upwind (FOU) approximation is then obtained via

$$
\phi_{FOU} = \begin{cases} 
\phi_0 & \text{if } \bar{m}_f \geq 0, \\
\phi_1 & \text{if } \bar{m}_f < 0,
\end{cases} \quad (3.23)
$$

using the cell centre values $\phi_0$ and $\phi_1$ on either side of the face. Furthermore, the second-order upwind approximation is computed as

$$
\phi_{SOU} = \begin{cases} 
\phi_{f,0} & \text{if } \bar{m}_f \geq 0, \\
\phi_{f,1} & \text{if } \bar{m}_f < 0,
\end{cases} \quad (3.24)
$$

using a linear interpolation of the face values

$$
\phi_{f,0} = \phi_0 + (\bar{x}_f - \bar{x}_0) \left( \nabla \phi \right)_{r,0}, \quad (3.25)
$$

$$
\phi_{f,1} = \phi_1 + (\bar{x}_f - \bar{x}_1) \left( \nabla \phi \right)_{r,1}, \quad (3.26)
$$

with the cell centre values $\phi_0$, $\phi_1$ and the gradients $\phi_{f,0}$, $\phi_{f,1}$ from either side. Finally, the central-differencing scheme computes the cell face value by linear interpolation between the neighbouring cell centre values as

$$
\phi_f = f \phi_0 + (1 - f) \phi_1, \quad (3.27)
$$

with a linear interpolation factor $f$ related to the mesh stretching ($f = 0.5$ for a uniform mesh).

- The diffusion term gives the net rate of increase of the variable $\phi$ due to diffusion and is discretised as

$$
D_f = \left( \Gamma \nabla \phi \cdot \bar{A} \right)_f = \Gamma_f \nabla \phi_f \cdot \bar{A} = \\
\Gamma_f \left\{ (\phi_f - \phi_0) \bar{\alpha} \cdot \bar{A} + \nabla \phi \cdot \bar{A} - \left[ \nabla \phi \cdot (\bar{x}_1 - \bar{x}_0) \right] \bar{\alpha} \cdot \bar{A} \right\}, \quad (3.28)
$$

with the face diffusivity $\Gamma$, the gradient of the fluid variable $\nabla \phi$ and the parameter

$$
\bar{\alpha} = \bar{A} / \left[ \bar{A} \cdot (\bar{x}_1 - \bar{x}_0) \right]. \quad (3.29)
$$

The gradient

$$
\nabla \phi = \frac{\nabla \phi_0 + \nabla \phi_1}{2} \quad (3.30)
$$

is computed as the mean from the two neighbouring cells.
Reynolds-Averaged Navier-Stokes Simulation. Initially, a Reynolds-averaged Navier-Stokes (RANS) simulation of the flow is conducted in order to obtain an estimate of the mean pressure field. In this approach, the governing Navier-Stokes equations of the flow are solved numerically in a time-averaged manner. This is achieved by applying Reynolds decomposition and splitting up flow velocity \( u_i (r', t) = \overline{u}_i (r', t) + u'_i (r', t) \) and pressure \( p (r', t) = \overline{p} (r', t) + p' (r', t) \) into its mean part, denoted by the overbar, and its fluctuating part, denoted by the prime. The time average of a flow variable \( \psi (r', t) \) is thus defined as

\[
\overline{\psi} (r', t) = \lim_{T \to \infty} \frac{1}{T} \int_{r}^{r+T} \psi dt,
\]

with the average of the fluctuating part vanishing as

\[
\overline{\psi'} (r', t) \equiv 0.
\]

When applying the Reynolds average as presented in Eq. (3.31), the density fluctuations are neglected. However, for the purpose of establishing an estimate of the mean flow properties and for flow cases of \( M \ll 1 \), as in the human upper airways, it is a fast alternative. The resulting set of mean flow equations then reads as

\[
\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i} (\rho \overline{u}_i) = 0,
\]

\[
\frac{\partial}{\partial t} (\rho \overline{u}_i \overline{u}_j) + \frac{\partial}{\partial x_i} (\rho \overline{u}_i \overline{u}_j) = -\frac{\partial p}{\partial x_j} + \frac{\partial}{\partial x_i} (\overline{\sigma}_{ij} + \sigma^R_{ij}).
\]

Here the \( \overline{\sigma}_{ij} \) is the mean viscous shear stress tensor given by the time-averaged net momentum flux through molecular motion at microscopic scales. The turbulent stress tensor \( \sigma^R_{ij} \), also known as Reynolds stress tensor, gives the time-averaged transport of momentum through velocity fluctuations at macroscopic scales. It is defined as

\[
\sigma^R_{ij} = -\rho \overline{u'_i u'_j}.
\]

In this initial RANS-based approach, the SST \( k-\omega \) eddy-viscosity model after Menter (1992) is applied.

Large Eddy Simulation. The chosen method of this work is large eddy simulation (LES), in order to compute the instantaneous disturbances. It resolves fluctuations of the flow variables in the large scales, which contain the majority of the turbulence energy. As opposed to RANS, where the underlying equations are derived from averaging, LES applies a filtering operation to the flow variable

\[
\tilde{\psi} (r', t) \equiv \int_V \psi (r_s, t) G (r, r_s) d\tilde{r}_s,
\]
by use of the filter function \( G(\vec{r}, \vec{r}_s) = G(\vec{r} - \vec{r}_s) \) that acts as a convolution kernel. Thus, the quantity can be decomposed into a filtered part, specified by the tilde and defined in Eq. (3.36), and a subfilter one, denoted by the \( s \):

\[
\tilde{\psi}(\vec{r}, t) = \tilde{\psi}(\vec{r}, t) + \psi_s(\vec{r}, t).
\] (3.37)

Unlike the Reynolds decomposition, filtered fluctuations arising from LES are not generally equal to zero as

\[
\tilde{\psi}_s \neq 0, \quad \tilde{\psi} \neq \tilde{\psi}.
\] (3.38)

For the time-dependent computations, the spatially filtered Navier-Stokes equations for the momentum of the fluid,

\[
\frac{\partial}{\partial t}(\rho \tilde{u}_i) + \frac{\partial}{\partial x_j}(\rho \tilde{u}_i \tilde{u}_j) = -\frac{\partial P}{\partial x_i} + \frac{\partial \sigma_{ij}}{\partial x_j} + \frac{\partial \sigma_{ij}^{SGS}}{\partial x_j}
\] (3.39)

with the shear stress tensor \( \sigma_{ij} \) and the subgrid stress tensor \( \sigma_{ij}^{SGS} \), are calculated for a numerical mesh size small enough to properly handle the complex geometry of the upper airways and to resolve the most important, energy containing flow structures. The subgrid scales of the flow, which are smaller than the numerical cell size, are low-pass filtered by the computational grid. Thus, the filter width is directly related to the spatial resolution of the mesh.

Chen & Gutmark (2014) were able to show that LES gives better estimates of the root mean square flow field than the Reynolds stress model (RSM). This is one of the reasons for using this approach for the computations performed in this study, as the prediction of near-field aerodynamic fluctuations is crucial for the prediction of acoustic sources. The underlying computational meshes are refined with regards to regions of high shear and large gradients as well as towards the boundaries of the domain. Grid spacing \( \Delta x \) and time step \( \Delta t \) are chosen to satisfy the Courant-Friedrichs-Levy (CFL) criterion. The convective Courant number \( CFL_c = u \Delta t / \Delta x \) and the acoustic one \( CFL_a = c_0 \Delta t / \Delta x \) are related via the Mach number

\[
\frac{CFL_c}{CFL_a} = \frac{u}{c_0} = M.
\] (3.40)

As the Mach number of the considered flow regime is small \((M \ll 1)\), the convective Courant number becomes the critical parameter for the choice of grid and time step sizes to capture all the aerodynamic fluctuations in the source region \((CFL_c \ll CFL_a)\).

The maximum time step of the computation is chosen as \( \Delta t_{\text{max}} = 2 \times 10^{-5} \text{s} \) (20 \( \mu \text{s} \)), sufficiently small to resolve the full Fourier spectrum of frequencies relevant for phonation. The fluctuations in velocity and pressure are computed in a coupled manner.
Turbulence Modelling. The behaviour of a flow is often characterised by the Reynolds number, which is defined as the ratio between inertial and viscous forces, or

\[ Re = \frac{u_0 d}{\nu}, \]  

(3.41)

based on the mean velocity magnitude \( u_0 \), the kinematic viscosity \( \nu \), and the characteristic length scale \( d \), which in the case of laryngeal flows is often taken as the instantaneous glottal diameter. Thus, the Reynolds number of the flow will reach a maximum at the glottal constriction. In general, flows related to the human phonation process are estimated to have a Reynolds number of \( O(100) \)–\( O(10000) \) (Mittal et al. 2013). For the simulations conducted in this work, the maximum Reynolds number is estimated as \( Re_{max} \sim O(3000) \). Although the laryngeal airflow is laminar in most cases, it becomes transitional in the glottal jet and its shear layers and can develop turbulence. However, since glottal flow is pulsating under voiced speech, the turbulent state of the flow is never fully developed. Nevertheless, it is ensured that all relevant scales of turbulence are resolved in this work, where relevant.

The eddies of large scales in a turbulent flow contain the most energy, which is provided by the mean flow. This energy is then transmitted to smaller and smaller scales until viscous stresses become important and the kinetic energy of the flow dissipates as heat. This process, the turbulence energy cascade, is happening in the inertial subrange, where scales are of the order of the Taylor microscale. It lies between the energy-containing range of the integral length scale and the dissipation range of the Kolmogorov length scale. At the limit of high Reynolds numbers, the turbulence in the flow can be assumed homogeneous and isotropic.

As hypothesised by Kolmogorov (1941), this means that any small scales of turbulence depend solely on kinematic viscosity \( \nu \) and viscous dissipation \( \varepsilon \). By using dimensional analysis and the known dimensions of these parameters \( [\eta] = \text{m}, [\nu] = \text{m}^2/\text{s}, [\varepsilon] = \text{m}^2/\text{s}^3 \), the Kolmogorov length scale can be derived as \( \eta = f(\nu, \varepsilon) = \nu^{2/3} \varepsilon^{1/3} \). Similarly, the associated time scale is \( t_\eta = (\nu/\varepsilon)^{1/2} \) and the velocity scale is \( u_\eta = (\nu \varepsilon)^{1/4} \). In the universal equilibrium range, a balance is assumed between the energy transferred to the large scales and the energy dissipated at the small scales, such that \( \partial E/\partial t \approx 0 \). Under these conditions, a model for the energy spectrum function is

\[ E(k) = C\varepsilon^{\frac{2}{3}} k^{-\frac{5}{3}} f_\eta(k\eta), \]  

(3.42)

for wavenumbers \( k \). The function \( f_\eta(k\eta) \) is unity in the inertial subrange, such that the classic Kolmogorov \(-5/3\) spectrum is recovered:

\[ E(k) = C\varepsilon^{\frac{2}{3}} k^{-\frac{5}{3}}. \]  

(3.43)

This spectral model has been initially derived in the wavenumber domain, but relations between the temporal and spatial fluctuations of turbulence have
been investigated early. More recently, Wilczek & Narita (2012) show through integration of the full wavenumber space, that a similar Eulerian frequency spectrum applies:

\[ E(\omega) = C_v \omega^{4/7} |\omega|^{-4/7}. \]  

(3.44)

The assumption of homogeneous and isotropic turbulence implies that the statistical distribution of small-scale turbulent eddies is independent from any boundaries or spatial directions. However, this is an ideal case and barely applies to realistic flow in confined scenarios such as biological flows.

The use of LES in this study implies that scales in the airflow are resolved down to the size of the applied volume cells. Therefore, it is crucial to match the spatial discretisation to the Taylor microscale, which is computed following an estimate by Pope (2001):

\[ \lambda_T \approx \left[ 15\nu \times 0.1d/(0.1u_{k,max}) \right]^{1/2}. \]  

(3.45)

It is based on the integral length scale \( l = 0.1d \), which is assumed to be one order of magnitude smaller than the characteristic length scale of the geometry in the region of interest, which is the glottal width \( d \). Furthermore, it includes the values for the kinematic viscosity of air of \( \nu \approx 1.79 \times 10^{-5} \) m\(^2\)/s and the maximum axial velocity \( u_{k,max} \). The approximation of Eq. (3.42) is used to design the numerical grids in this work with cell dimensions adapted to it. For modelling of the subgrid scales, the wall-adapting local eddy viscosity (WALE) model, after Nicoud & Ducros (1999), is applied in this work. It models the subgrid scale viscosity as

\[ \nu_t = \Delta^2 S_w, \]  

(3.46)

with a subgrid characteristic length scale \( \Delta = C_w V^{1/3} \), typically related to the cell volume \( V \) via a model coefficient \( C_w \), and the deformation parameter \( S_w \), which is a function of the resolved velocity field and the velocity gradient tensor.

3.3. Sound Sources and Acoustic Analogies

On the one hand, compressible large eddy simulations (LES), or even direct numerical simulations (DNS) have been used in the past for directly computing acoustic pressure levels generated and propagated by flows. On the other hand, acoustic analogies are regularly applied through hybrid approaches to obtain the acoustic signal generated by aerodynamic sources. Acoustic analogies are generally achieved through a reformulation of the governing equations of fluid motion in the region of aerodynamic fluctuations into the form of a wave equation. The inhomogeneity of the wave equation can then be interpreted as an acoustic source field. This is especially relevant within the field of speech science, in order to both quantify and localise the important acoustic source types produced through the sound-generating mechanisms discussed in Sec. 2.1.
Furthermore, the nature of the pressure perturbations is dependent on the domain. In the source region (i.e. near field) they are due to flow fluctuations and turbulence, convected with the flow velocity, while in the propagation region (i.e. far field) they are caused by acoustic fluctuations that are irrotational, moving with the speed of sound.

3.3.1. Lighthill’s Acoustic Analogy

Lighthill (1952) was the first to draw the analogy of acoustic wave propagation to fluid mechanics for the application of jets into free stream. The Lighthill analogy suggests a direct causal connection between the aerodynamic pressure fluctuations in a near field and the acoustics in the far field. It considers the flow-induced perturbations in the near field as source for an inhomogeneous wave equation obtained through reformulation of the governing equation system of fluid mechanics, in this case the Navier-Stokes equation for compressible viscous flow (3.10):

\[
\left( \frac{1}{c_0^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) [c_0^2 (\rho - \rho_0)] = \frac{\partial^2 T_{ij}}{\partial x_i \partial x_j} \quad (3.47)
\]

Eq. (3.47) is the Lighthill equation, in terms of the density fluctuation \( \rho - \rho_0 = \rho' \) and the double divergence of the Lighthill stress tensor

\[
T_{ij} = \rho u_i u_j + \delta_{ij} (p_0 + p' - c_0 \rho') - \sigma_{ij}, \quad (3.48)
\]

with the viscous stress tensor \( \sigma_{ij} \). \( T_{ij} \) is the strength per unit volume of the distribution of acoustic sources. Apart from the generation of sound, it also accounts for convection by the flow and refraction as a result of sound speed variations ( Howe 1998). The expression \( c_0^2 (\rho - \rho_0) \) tends to the pressure fluctuation \( p' = p - p_0 \) as \( |\mathbf{v}| \to \infty \) in the fluid outside the source region.

3.3.2. Flowcs Williams-Hawkings Equation

Lighthill’s equation was consequently generalised by Williams & Hawkings (1969) in order to include the interaction of the flow with solid surfaces in the acoustic source region. The general Flowcs Williams-Hawkings (FWH) equation is

\[
\left( \frac{\partial^2}{\partial t^2} - c_0^2 \nabla^2 \right) [H(S) \rho'] = \frac{\partial^2}{\partial y_i \partial y_j} \left[ T_{ij} H(S) \right] - \frac{\partial}{\partial y_i} \left\{ p \delta_{ij} - \sigma_{ij} + \rho u_i (u_j - v_j) \frac{\partial S}{\partial y_j} \delta(S) \right\} + \frac{\partial}{\partial t} \left\{ \rho_0 v_i + \rho (u_i - v_i) \frac{\partial S}{\partial y_i} \delta(S) \right\}, \quad (3.49)
\]

where \( H(S) \) is the Heaviside function, \( \delta(S) \) is the Dirac function, and \( S(y,t) \) defines the geometry of the wall boundary with \( S > 0 \) inside the flow region and \( S < 0 \) outside. Eq. (3.49) contains additional terms for the relative
motion between the fluid of velocity \( u_j \) and the solid boundaries of velocity \( v_j \). However, its original solution does only apply to unconfined flow scenarios in free stream, which allow for an unperturbed propagation of the generated acoustic waves from the near field to the far field. This is not the case in the considered airway geometries of this study. Nevertheless, a duct-like geometry permits a few assumptions for the solution of the FWH equation:

- Propagated waves can be assumed planar,
- The propagation direction can be assumed one-dimensional.

The chosen Green's function fitting with these assumptions is

\[
g(x_1,t|y_1,\tau) = \frac{c_0}{2} H \left( \tau - \frac{|x_1 - y_1|}{c_0} \right)
\]

defining a pulse being released at the source location \( y \) at time \( \tau \) and detected at the observation location \( x \) at time \( t \) (Hofmans 1998). The resulting integral expression gives the pressure fluctuation field originating from the source region and satisfies the boundary conditions of the domain:

\[
p'(x,t) = \frac{1}{2S_0c_0} \frac{\partial}{\partial t} \int_V \left[ \rho u_1^2 - \sigma_{11} \right]_{\tau} dS \\
\quad - \frac{1}{2S_0} \int_{S_{wall}} \left[ p' \delta_{ij} - \sigma_{ij} \right]_{\tau} \cdot n_j \times \text{sign}(x_1 - y_1) dS \\
\quad - \frac{1}{2S_0} \int_{S_{wall}} \left[ \rho_0 c_0 v_j \right]_{\tau} \cdot n_j dS \\
\quad + \frac{1}{2S_0} \int_{S_{inlet}} \left[ p' + \rho_0 c_0 u_1 \right]_{\tau} dS \\
\quad + \frac{1}{2S_0} \int_{S_{inlet}} \left[ p' - \rho_0 c_0 u_1 \right]_{\tau} dS,
\]

with the cross-sectional area of the duct \( S_0 \), the source region \( V \), the unit normal vector \( n_j \), and the retarded time \( t^* = t - \frac{|x_1 - y_1|}{c_0} \). The terms on the right hand side of Eq. (3.51) can be attributed to each source type of aerodynamic sound in the near field. These terms further simplify or cancel out considering the particular set of boundary conditions and flow properties present. Term
(I) can be identified as a quadrupole source due to the turbulent flow inside the domain. Term (II) is a dipole source due to the unsteady forces imposed by the solid surfaces on the fluid. Term (III) is a monopole source related to the motion of surfaces. The terms (IV) and (V) are monopole sources related to reflections at the inlet and outlet boundaries. These terms can be neglected in case of acoustically non-reflecting boundary conditions as applied in this study. Additionally, the viscous stress components $\sigma_{ij}$ in terms (I) and (II) can be ignored for mean inflow velocities under the physiological conditions of the human airways (Zhang et al. 2002b).

3.3.3. Source Types of Phonation

The acoustic analogy presented in Sec. 3.3.2 can be applied to the voice generation process, allowing to assess the strength of the various source types.

For aerodynamically generated sound, two main sources can be identified: the velocity variations associated with the developed vortical flow structures in a turbulent flow scenario and the fluctuations of pressure on a surface. The latter are caused by the unsteady flow and vortices impacting on the boundaries of a rigid or elastic body. Therefore, a turbulent flow field can be seen as a distribution of acoustic quadrupole sources, while surface perturbations are typically dipole sources.

First, McGowan (1988) took an aeroacoustic approach to phonation by solving for the present sources: He identified a monopole source linked to the oscillation frequency of the vocal folds, leading to a localised volume modulation and displacement of air around the solid glottal surfaces. Furthermore, two types of dipole sources were found: one produced by vorticity-velocity interaction and the other through an exchange of energy with the conversion of irrotational to rotational flow. For regular phonation, it was found that the dipole source term has the greatest strength. The monopole source was found to increase with the fundamental frequency of oscillations and becomes dominant at about 400 Hz (Zhao et al. 2002; Zhang et al. 2002a). Finally, Hirschberg (1992) used Lighthill’s acoustic analogy (cf. Sec. 3.3.1) to prove that quadrupole sources are also contributing to the sound of voiced speech. The broadband sound component of speech is caused by these quadrupole sources as a result of turbulence kinetic energy fluctuations. It has been shown experimentally to increase significantly at high flow rates (Zhang et al. 2004). Additionally, quadrupole sources are considered the dominant acoustic source types in case of unvoiced speech, such as fricatives and whispering (Hirschberg 1992; Howe & McGowan 2005).

3.4. Rigid Glottal Model

The initial geometry for verification of the applied numerical method, as well as first studies of the flow structures generated at the glottis are based on
the so-called M5 model for the vocal folds, which was introduced by Scherer et al. (2001). It is based on average values of the anatomical dimensions of

![Diagram of vocal fold model](image)

**Figure 3.1.** Geometrical dimensions of the M5 vocal fold model in the divergent configuration as introduced by Scherer et al. (2001).

the glottal constriction (cf. Tab. 3.1) and was used for the experimental study of the pressure distribution at the airway walls. The vocal fold surface design equations for the geometry depicted in Fig. 3.1 are

\[
R_\phi = \frac{R_0}{1 - \sin \left( \frac{\psi}{2} \right)}, \quad R_L = \frac{T}{2}
\]  

(3.52)

\[
B = \frac{\sqrt{2} R_\psi}{\sqrt{1 + \sin \left( \frac{\psi}{2} \right)}} = \frac{R_0 \sec \left( \frac{\psi}{2} \right)}{\sqrt{\frac{1}{2} \left[ 1 - \sin \left( \frac{\psi}{2} \right) \right]}}
\]  

(3.53)
\[ B = \frac{\sqrt{2}R_\psi}{\sqrt{1 + \sin\left(\frac{\psi}{2}\right)}} \]

\[ = \left[ T - R_0 - R_L \sin\left(\frac{\psi}{2}\right) \right] \sec\left(\frac{\psi}{2}\right), \quad (3.54) \]

\[ Q_2 = R_L \sin\left(\frac{\psi}{2}\right), \quad Q_3 = Q_1 \cos\left(\frac{\psi}{2}\right), \quad (3.55) \]

\[ Q_4 = R_6, \quad Q_5 = R_L \sin\left(50^\circ\right), \quad (3.56) \]

with human values for the curvature radius \( R_0 = 0.0987 \text{ cm}, \) the axial length \( T = 0.3 \text{ cm}, \) and the glottal angle \(-40^\circ \leq \psi \leq 40^\circ. \) After initial validation

<table>
<thead>
<tr>
<th>Laryngeal airway variable</th>
<th>Physical values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal length (cm)</td>
<td>1.2</td>
</tr>
<tr>
<td>Glottal diameter (cm)</td>
<td>0.04–0.08</td>
</tr>
<tr>
<td>Glottal thickness (cm)</td>
<td>0.3</td>
</tr>
<tr>
<td>Glottal entrance angle (°)</td>
<td>-40–40</td>
</tr>
<tr>
<td>Subglottal lateral diameter (cm)</td>
<td>2.0</td>
</tr>
<tr>
<td>Subglottal A-P diameter (cm)</td>
<td>2.0</td>
</tr>
<tr>
<td>Supraglottal lateral diameter (cm)</td>
<td>2.0</td>
</tr>
<tr>
<td>Supraglottal A-P diameter (cm)</td>
<td>2.0</td>
</tr>
</tbody>
</table>

of the numerical method of this study (cf. Sec. 3.6), the vestibular folds are added to the model to account for realistic confinement of the glottal jet. The computational mesh of the rigid glottal model is unstructured and consists of polyhedral elements. It is locally refined at the constriction of true and false vocal folds to capture the flow separation points and vortex shedding, as shown in Fig. 3.2. Three prism layers are applied towards the solid boundaries and the mesh is extruded from the in- and outlet with increasing cell base size to avoid the reflection of pressure fluctuations. The boundaries of the domain are modelled as static walls with no-slip condition. A constant flow velocity is imposed at the inlet, while the stagnation pressure at the outlet is atmospheric.

3.5. MRI-Based Vocal Tract Model

According to Arnela et al. (2016), only detailed, three-dimensional vocal tract representations are able to correctly predict the high-frequency modes occurring
Figure 3.2. Geometry of the rigid glottal model including both true vocal folds (TVF) and false vocal folds (FVF) (a) and a magnified plane section of the unstructured mesh at the glottal constriction (b).

in the voice spectrum. In order to obtain realistic models of the vocal tract for the pronunciations of different vowels, geometries based on MRI data of a 30 year-old, healthy male subject are applied. The MRI recordings were performed by the Speech Modelling Group at the Department of Mathematics and Systems Analysis, Aalto University, and capture the anatomy of the upper airway cavities during phonation with the subject in supine position (cf. Fig. 1.6). The numerical meshes based on the MRI geometries are unstructured and consist of the fluid domain ranging from the airway channels to a far-field domain terminated by acoustically non-reflecting boundaries. It is shown for the vowel [a] in Fig. 3.3. The grids for the different vowel articulations used in this study have a resolution based on the results of a grid convergence study (cf. Sec. 3.6.1) with a final cell count of approximately 5.85 \cdot 10^6. The airflow is considered as compressible according to Sec. 3.2.2 and the surrounding tract is assumed as rigid.

3.5.1. Glottal Waveform

In parts of this study, the inlet boundary condition is given as a periodic function of the volumetric flow rate simulating the flow modulation of the oscillating vocal folds (cf. Papers 1-3). The Rosenberg waveform \( f_R(t) = f_p(t) + f_n(t) \) describes the approximate shape of the glottal pulse produced through the regular opening and closing of the vocal folds (Rosenberg 1971). It consists of the function

\[
f_p(t) = \frac{A_0}{2} \left[ 1 - \cos \left( \frac{t}{t_p} \pi \right) \right] \quad 0 \leq t \leq t_p, \tag{3.57}
\]
for its positive slope and the function

$$f_n(t) = A_0 \cos \left( \frac{t - t_p \pi}{t_n} \right) \quad t_p \leq t \leq t_p + t_n, \quad (3.58)$$

for its negative slope, with the pulse amplitude $A_0$. The relative duration of the glottal opening $t_p$ with respect to the total pulse length $t_{tot} = t_p + t_n$ is referred to in the following as glottal opening quotient. In the example given in Fig. 3.4 the glottal pulse shows a fundamental frequency of 120 Hz, characteristic for a male voice.

Free stream boundary conditions are imposed at the outlet to the acoustic far field with the fluid characteristics being specified by Mach number, atmospheric pressure, and room temperature.

3.5.2. Quasi-Steady Approximation

One approach to investigate the supraglottal flow field and the aerodynamically generated acoustic sources is to consider the flow through static laryngeal models of specific glottal angles. This method, referred to as quasi-steady approximation has been shown to be valid by Zhang et al. (2002b) for application in phonation and speech acoustics. Thus, the process of speech production can
also be approximated by simulations with steady inflow through static laryngeal models of geometric dimensions related to certain instances of the glottal cycle, as shown in Fig. 1.3.

In this work, this approach is taken to study the sound generation of phonation by aerodynamic mechanisms and without the contribution of the harmonic, volume-modulating monopole source of at the vocal folds (cf. Paper 4). A constant stagnation pressure equal to a subglottal pressure of $P_0 = P_e \approx 588.399 \text{ Pa} \ (6 \text{ cmH}_2\text{O})$ is imposed at the subglottal inlet of the airways upstream of the glottal constriction. A pressure outlet at atmospheric pressure is chosen for the acoustic far field boundaries.

3.6. Verification and Validation

3.6.1. Grid Convergence Study

The appropriate resolution of the unstructured mesh is chosen following a grid convergence study including five grids of decreasing base size of the cells and increasing cell count. The procedure follows the recommendations given by Celik et al. (2008). A grid refinement factor of $r = h_{\text{coarse}}/h_{\text{fine}} \geq 1.3$ is chosen for the study, with an initial value of $r = 2$ being reduced to $r = 1.3$ for the finest grids. Tab. 3.2 gives the base size of the polyhedral cells, the chosen time step of the simulation, and the approximate cell count. During the grid convergence study, both base size and time step are reduced by the same factor to avoid interaction of spatial and temporal discretisation errors. The values to analyse are the first and second statistical moments (mean and variance) of the important flow variables, in this case velocity and pressure. They are extracted
Table 3.2. Properties of the numerical grids and simulations applied for the convergence study.

<table>
<thead>
<tr>
<th>Grid no.</th>
<th>Base size (m)</th>
<th>Time step (s)</th>
<th>Cell count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$5 \cdot 10^{-3}$</td>
<td>$1 \cdot 10^{-4}$</td>
<td>$\sim 0.1 \cdot 10^6$</td>
</tr>
<tr>
<td>2</td>
<td>$2.5 \cdot 10^{-3}$</td>
<td>$0.5 \cdot 10^{-4}$</td>
<td>$\sim 0.6 \cdot 10^6$</td>
</tr>
<tr>
<td>3</td>
<td>$1.25 \cdot 10^{-3}$</td>
<td>$0.25 \cdot 10^{-4}$</td>
<td>$\sim 0.9 \cdot 10^6$</td>
</tr>
<tr>
<td>4</td>
<td>$0.9615 \cdot 10^{-3}$</td>
<td>$0.192 \cdot 10^{-4}$</td>
<td>$\sim 1.7 \cdot 10^6$</td>
</tr>
<tr>
<td>5</td>
<td>$0.7396 \cdot 10^{-3}$</td>
<td>$0.148 \cdot 10^{-4}$</td>
<td>$\sim 3.8 \cdot 10^6$</td>
</tr>
</tbody>
</table>

![Figure 3.5. Placement of line probes in the axial and lateral direction at the glottis, shown in the sagittal plane section of the numerical mesh.](image)

From the computation via line probes near the glottal constriction and in the supraglottal region, as shown in Fig. 3.5. The mean values of both velocity and pressure converge well within the considered range of mesh sizes (cf. Fig. 3.6(a) and Fig. 3.7(a)). The same applies for the variance of these quantities along the centreline $LC$ (cf. Fig. 3.6(b) and Fig. 3.7(b)). Additionally, time-averages of velocity and pressure are shown for the lateral line probes $L1$, $L2$, and $L3$ in Fig. 3.8. Since the precise computation of pressure fluctuations is crucial for direct aeroacoustics calculations, also the variance of pressure is compared for these line probes in perpendicular direction to the flow (cf. Fig. 3.9).

The discretisation errors are computed for the investigated base cell sizes $h_i < h_j < h_k$ of the grids with numbers $i < j < k$ by first calculating the order
Figure 3.6. Mean (a) and variance of velocity (b) along the axial centreline $LC$ of the glottal jet for the three finest grid resolutions of Tab. 3.2.

Figure 3.7. Mean (a) and variance of pressure (b) along the axial centreline $LC$ of the glottal jet for the three finest grid resolutions of Tab. 3.2.

$p$ of the method

$$p = \frac{1}{\ln(r_{ji})} \ln \left( \frac{\varepsilon_{kj}}{\varepsilon_{ji}} \right) + q(p),$$  \hspace{1cm} (3.59)

with

$$q(p) = \ln \left( \frac{r_{ji}^p - s}{r_{kj}^p - s} \right),$$  \hspace{1cm} (3.60)

$$s = 1 \cdot \text{sgn} \left( \frac{\varepsilon_{kj}}{\varepsilon_{ji}} \right).$$  \hspace{1cm} (3.61)
Figure 3.8. Time-averaged velocity (a) and pressure (b) along the line probes L1, L2, and L3 shown in Fig. 3.5.

Figure 3.9. Variance of pressure along the line probes L1, L2, and L3 shown in Fig. 3.5.

The absolute errors are computed as

$$
\varepsilon_{ij} = \phi_j - \phi_i, \quad \varepsilon_{kj} = \phi_k - \phi_j,
$$

(3.62)

with the numerical solution $\phi_i$ for the i-th grid. The extrapolated relative error is obtained as

$$
\varepsilon_{ij}^a = \left| \frac{\phi_i - \phi_j}{\phi_i} \right|,
$$

(3.63)
and further used for the calculation of the grid convergence index (GCI):

\[ GCI_{\text{fine}}^{ji} = \frac{1.25 \epsilon^{ji}_{\text{fine}}}{r_{ji} - 1}. \]  

The computed discretisation errors using Eq. (3.64) for the two finest grids of number 4 and 5 (cf. Tab. 3.2) are presented in Fig. 3.10 for the mean fields of the flow variables evaluated at the centreline probe \( LC \). It can be

![Figure 3.10. Discretisation errors for the mean values of pressure (a) and velocity (b).](image)

recognised in Fig. 3.10 that the errors are small and close to zero over a large distance. However, uncertainty in the flow velocity becomes larger before the glottal entrance and after the exit. Therefore, based on the convergence study of the mean values and the associated discretisation errors, the cell sizes of grid no. 5 of Tab. 3.2 are chosen for further use in this work with additional local mesh refinement in the sensitive domains surrounding the glottis. Thus, the final meshes for the vocal tract geometries have an approximate cell count of \( 5 \cdot 10^6 \) or higher, depending on the extension of the acoustic far-field domain. The chosen mesh sizes are well above the minimum requirement of 40 points per wavelength for the highest frequencies occurring through phonation and are thus suited for extracting the relevant acoustic information.

### 3.6.2. Glottal Pressure Distribution

A validation study is carried out to assess the performance of the solver towards accurate prediction of the pressure at the airway walls. Thus, the rigid glottal model, as described in Sec. 3.4, is subjected to particular subglottal pressures at the inlet to compare against experimental results under the same conditions. Fig. 3.11 shows the pressure profiles for the three investigated cases. The differentiation between flow and non-flow wall is due to a deflection of the
glottal jet towards one of the airway sides, which leads to asymmetry in the downstream region and is further explained in Sec. 5.2.

3.6.3. Acoustic Formant Data

Finally, acoustic validation of the applied numerical approach is performed by comparison of the obtained dominant tonalities of the vocal tract volumes from the unsteady LES simulations with the sound data extracted simultaneously with the MRI measurements by the Speech Modelling Group at the Department of Mathematics and Systems Analysis at Aalto University. As shown in Tab. 3.3, good overall agreement is reached, considering the considerable uncertainties of both numerical and experimental methods for formant extraction, which are further discussed in Paper 3.
Table 3.3. Numerical and experimental formant data extracted through LES simulations of the compressible airflow and through non-metallic waveguides during MRI recording of the vowel vocalisation.

<table>
<thead>
<tr>
<th></th>
<th>Formant $F_1$ (Hz)</th>
<th>Formant $F_2$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>numerical</td>
<td>experimental</td>
</tr>
<tr>
<td>Vowel [a]</td>
<td>480</td>
<td>651</td>
</tr>
<tr>
<td>Vowel [e]</td>
<td>484</td>
<td>542</td>
</tr>
<tr>
<td>Vowel [i]</td>
<td>242</td>
<td>247</td>
</tr>
<tr>
<td>Vowel [o]</td>
<td>359</td>
<td>398</td>
</tr>
<tr>
<td>Vowel [u]</td>
<td>240</td>
<td>306</td>
</tr>
</tbody>
</table>
CHAPTER 4

Modelling of Airway Obstructions

Soft tissues, like the ligaments and mucosa in the human upper airways, show mechanical properties that depend strongly on factors such as age, gender, and health status, but also on temperature, osmotic pressure, and strain rate (Holzapfel 2001).

Due to the wide range of effects and phenomena that can be attributed to biological and biomechanical fluid-structure interactions (FSI), it remains a highly relevant field of research in the area of physiological fluid mechanics. An important factor is the flexibility of the airway that leads to displacement of the soft tissue and thus significantly alters the flow conditions and acoustics. This challenge has been partially addressed by the quasi-steady approximation, using the steady flow fields inside static geometries at different stages of the deformation to make predictions about the dynamic behaviour (Zhang et al. 2002b; Lucey et al. 2010).

However, ultimately time-dependent deformations need to be taken into account by modelling the interaction between the airflow and the elastic walls (cf. Sec. 4.1).

Although taking a minor part in the context of this thesis, a simplified geometry of the pharyngeal airway with an elastic object resembling the soft palate is considered (cf. Sec. 4.2).

4.1. Fluid-Structure Interaction

In this work, FSI is simulated in an explicit two-way coupling approach. The absolute and static pressure as well as the wall shear stress fields at the interfaces are used for computing the overall stress imposed on the structure at each time step, while the displacement fields are fed back in order to rearrange the grid via mesh morphing at each inner iteration.

For the purpose of this study and for the considered degree of tissue displacement, a linear isotropic elastic stress-strain relationship, valid for small strains, is employed. In this approach, the strain field of the solid is directly related to the gradient of the displacement field. The stress field is linked to
the strain field via the constitutive law

$$
\sigma_{ij} = 2G \left[ \varepsilon_{ij} - \delta_{ij} \frac{1}{3} \sum_{i=1}^{3} \varepsilon_{ii} \right] + \delta_{ij} K \sum_{i=1}^{3} (\varepsilon_{ii} - \varepsilon_{ii}^{T}) , \quad (4.1)
$$

expressing the stress tensor $\sigma_{ij}$ as a linear function of the strain tensor $\varepsilon_{ij}$ via the shear modulus $G$, the bulk modulus $K$, and the isotropic thermal strain tensor $\varepsilon_{ij}^{T}$. This relationship can also be expressed in terms of Young’s modulus $E$ and Poisson’s ratio $-\frac{dc_{\text{trans}}}{dc_{\text{axial}}}$ as

$$
E (\varepsilon_{ij} - \varepsilon_{ij}^{T}) = \left( 1 - \frac{dc_{\text{trans}}}{dc_{\text{axial}}} \right) \sigma_{ij} + \delta_{ij} \frac{dc_{\text{trans}}}{dc_{\text{axial}}} \sum_{i=1}^{3} \sigma_{ii} , \quad (4.2)
$$

with

$$
G = \frac{E}{2 \left( 1 - \frac{dc_{\text{trans}}}{dc_{\text{axial}}} \right)} , \quad (4.3)
$$

$$
K = \frac{E}{3 \left( 1 + 2 \frac{dc_{\text{trans}}}{dc_{\text{axial}}} \right)} . \quad (4.4)
$$

For the structural mechanics a finite element model is used to account for the deformation of the tissue as response to the shear stress and surface pressure imposed by the fluid on the boundary walls. This requires a simultaneous solution of both fluid and wall properties for a coupled FSI simulation. The numerical approach applied in this project has been validated for an extensive benchmark case (Turek & Hron 2006; Turek et al. 2011).

The acoustic signal of the vibrational response of the elastic parts in this study is assessed both via direct unsteady, compressible simulations and the acoustic analogy of the FWH equation (cf. Sec. 3.3.2). The integration domain includes both moving and static walls and the formulation takes into account excited, vibrating surfaces. Thus, a characterization of sources is pursued for the treated geometry to quantify the contribution of dipoles and monopoles to the total acoustic pressure. This is achieved by solving the individual source terms of the FWH equation in integral form (cf. Eq. (3.51)).

### 4.2. Elastic Element in Channel Flow

In order to perform a numerical analysis of the effects of vibrations of a flexible, beam-like structure in channel flow, the two-dimensional Navier-Stokes equations are solved for the flow around the model shown in Fig. 4.1 by means of the finite volume solver described in Sec. 3.2.2.

The model consists of an elastic element that is attached to a rigid circular cylinder. A similar setup, using a flexible filament attached to a circular cylinder in crossflow has been used by Bagheri et al. (2012) and Lācis et al. (2014) to numerically and experimentally study the effect of symmetry-breaking of passive, nature-inspired appendages.
In the course of this study, the aerodynamics and -acoustics of the system are investigated with regards to three parameters, all of which are indicated in Fig. 4.1: the Reynolds number of the flow $Re$, as well as length $l$ and elastic modulus $E$ of the flexible element. The length is also given in terms of the non-dimensional parameter $l^* = l/d$, where $d$ is the cylinder diameter, which is kept constant. The streamwise velocity at the inlet is specified as a parabolic, fully developed, laminar profile

$$u(0, y) = 1.5\bar{u} \frac{y(h-y)}{(\frac{h}{2})^2}, \quad (4.5)$$

modelling a conventional Poiseuille flow inside the pipe with the mean inflow velocity $\bar{u}$, the normal coordinate $y$, and the height of the channel $h$. The Reynolds number is defined as $Re = \bar{u} \cdot d/\nu$, with the mean inflow velocity $\bar{u}$, the obstacle diameter $d$, and the kinematic viscosity $\nu$ of air at room temperature, and is adjusted by variation of the mean inflow velocity at the inlet as introduced in Eq. (4.5). The value for $d$ is the maximum diameter of the obstacle. It is chosen to be one fourth of the channel height $h$, while the length of the flexible element $l$ is set to be 3.5 times the obstacle diameter $d$. The element of specified Poisson’s ratio $-d_{trans}/d_{axial} = 0.45$ and elastic modulus $E$ is compliant and deforms according to lift and drag forces exerted by the flow. Although the Mach number $M$ of the investigated cases of this study is well below 0.3, the flow is computed as compressible in order to capture any aeroacoustic effects.
CHAPTER 5

Phonation and Glottal Aerodynamics

This chapter describes selected findings of this work concerning the flow and acoustics related to phonation. It presents the important acoustic properties and resonances related to realistic upper airway geometries (cf. Sec. 5.1). Furthermore, the aerodynamic effects and coherent structures observed at the level of the larynx and epilarynx are examined (cf. Sec. 5.2).

The compressible flow solution obtained with the large eddy simulation (LES) approach outlined in Sec. 3.2.2 is used to characterize the fluid dynamics as well as to predict the sound generating sources according to Sec. 3.3, i.e. due to velocity variations in the turbulent flow or due to the unsteady pressure loads at the airway walls.

5.1. Generation of Vowel Sounds

The generation of different sounds and vowels during speech is caused predominately by the distribution of the formant frequencies influenced by the shape of the upper airways (cf. Sec. 2.1). The first formants $F_1$ and $F_2$ are characteristic for the spoken sound and can be recovered in the far-field acoustic spectra.

Using an intermittent inlet boundary condition for the volumetric flow rate as detailed in Sec. 3.5.1, the sound generation through voiced speech is simulated. Fig. 5.1(a) shows the propagation of acoustic waves from the surface model of the airway configuration of vowel [a] into the far field. The sound spectrum is extracted for a probe point $P$ which has a distance of approximately 8.5 cm from the mouth opening. The resulting spectrum is shown in Fig. 5.1(b) together with a spectral envelope and a graph of local maxima. The first formants $F_1$ and $F_2$ are indicated and compare well to the experimental values obtained simultaneously during the magnetic resonance imaging (MRI). The modes at the fundamental frequency $f_0$, as well as at the formants $F_1$ and $F_2$, which are extracted from the point-probe Fourier spectrum shown in Fig. 5.1(b), are further characterised by spatial cross-correlations. Fig. 5.2 gives space-time diagrams of the pressure fluctuations along a line probe to the far field. The slope of the disturbances is indicated and gives their propagation velocity as the speed of sound. Furthermore, temporal cross-correlations between near- and far-field probes indicate periods between the correlation peaks.
Figure 5.1. Acoustic pressure in the far field (a) and resulting Fourier spectrum of the sound pressure (b) for the vocal tract geometry of the vowel [a]. The first formants $F_1$ and $F_2$ are marked and compared to the Helmholtz resonance frequencies and experimental values obtained by the Speech Modelling Group at the Department of Mathematics and Systems Analysis, Aalto University.

Figure 5.2. Spatial cross-correlation in the vocal tract along a centerline from the oral cavity to the far field for the modes at frequencies $f_0 = 120 \text{ Hz}$, $F_1 = 480 \text{ Hz}$, and $F_2 = 1166 \text{ Hz}$ of the vowel [a]. The white dotted line indicates the slope for the speed of sound $c_0 = 343.21 \text{ m/s}$.

that correspond to the dominant modes $f_0$, $F_1$, and $F_2$. These correlation amplitudes are visible for all considered near-field probes in Fig. 5.3.
Figure 5.3. Temporal cross-correlation in absolute terms (a) and as correlation coefficient (b) of the pressure signals at the near-field probes $P_1$, $P_2$, $P_3$, and $P_4$ with respect to that at the far-field probe. The position of the probes is given in Fig. 5.9.

5.1.1. Cavity-Association of Formants

By applying Fourier transformation to the surface pressure fluctuations at the walls, the maxima of the sound pressure can be associated with different domains of the vocal tract. Thus, areas of resonance are identified and connected with the dominant frequencies of each vowel. The results for the vowel sounds [i], [e], [a], [o], and [u] are shown in Fig. 5.4. It can be seen that the first formant $F_1$ of high vowels such as [i] and [e], tends to be amplified in the laryngopharynx and oropharynx. For low vowels, such as [a] and [o], upper cavities, such as the oral, or even parts of the nasal cavity, are involved as resonators of the formant frequencies $F_1$ and $F_2$. The attributes low and high are referring here to the positioning of the tongue (Stevens 2000).

5.2. Glottal Flow and Supraglottal Flow Structures

As a result of the difference between the lung pressure in the subglottal region and the near-atmospheric pressure in the supraglottal region, flow is established through the opening of the vocal folds. Due to the variation of the glottal diameter, the maximum flow velocity is unsteady. Using the periodic volume velocity boundary condition of the Rosenberg glottal waveform explained in Sec. 3.5.1, the supraglottal flow fields are computed for various vowels. Fig. 5.5 shows that there is a clear difference in the flow field between the low vowel [a] and the high vowel [i]. Due to constriction by the tongue at high vowels, the flow is accelerated and the total upper airway volume is separated into
(a) Vowel [i]  (b) Vowel [e]  (c) Vowel [a]  (d) Vowel [o]  (e) Vowel [u]

\[ F_1 = 242 \text{ Hz} \quad F_1 = 484 \text{ Hz} \quad F_1 = 480 \text{ Hz} \quad F_1 = 359 \text{ Hz} \quad F_1 = 240 \text{ Hz} \]

\[ F_2 = 1900 \text{ Hz} \quad F_2 = 2037 \text{ Hz} \quad F_2 = 1156 \text{ Hz} \quad F_2 = 965 \text{ Hz} \quad F_2 = 940 \text{ Hz} \]

**Figure 5.4.** Sound pressure levels of the acoustic pressure fluctuations at the discrete formant frequencies \( F_1 \) and \( F_2 \) of the vowels [i], [e], [a], [o], and [u] in (a)-(e). The sound pressure level at the surface of the airway walls has been normalised by its maximum for each of the vocal tract cases.

The subvolumes resonating at particular frequencies relating to the formants (cf. Fig. 5.4).

5.2.1. **Glottal Jet Dynamics**

The intermittent flow through the vocal folds leads to the emergence of a jet that is pulsed by the periodic oscillation of the vocal folds. The quasi-steady approximation allows the investigation of the flow field and associated acoustics at instances of particular glottal shapes (cf. Sec. 3.5.2). Fig. 5.6 shows the glottal jet for two different opening diameters \( d_{VP} \) of the true vocal folds for the rigid glottal model, which includes both true and false vocal folds, as introduced in Sec. 3.4. Despite the symmetric laryngeal geometry, the jet is deviated towards one side of the channel. This deflection is found to be mostly due to shear layer instabilities in the area where the glottal jet detaches from the vocal fold surfaces (Zheng et al. 2011; Mittal et al. 2013).

If the Reynolds number is very low (e.g., \( Re \sim O(100) \)), the fluid flow will be stable and perfectly symmetric, leading to symmetric recirculation regions downstream. As the Reynolds number is increasing, a bifurcation of the
solution may occur and the flow attaches to one wall or the other. Due to this asymmetry in the initial glottal jet in the glottis, an asymmetry in the supraglottal region will develop with accompanying asymmetric recirculation regions. Furthermore, the formation of large-scale asymmetric flow structures and recirculation zones in the supraglottal region are likely to contribute to this effect and lead to a stabilisation of the jet towards one direction (Kniesburges et al. 2013).

The so-called Coanda effect is probably of minor importance and the use of the term in the field of phonation has been questioned (Mittal et al. 2013). The Coanda effect is the phenomenon of unconfined jet flow attaching to a close surface and following its curvature due to a favourable pressure gradient between the surrounding pressure and the surface pressure (Coanda 1936). Fig. 5.6 demonstrates that the glottal jet deflection becomes stronger at smaller distance $d_{YF}$, making it attach directly to the true vocal folds instead of the false vocal folds. The side to which the jet deflects is random and depends on physical instabilities causing the further development of supraglottal vertices. During one full glottal cycle, the jet might change orientation arbitrarily and switch from one side to the other, introducing additional surface pressure fluctuations. Furthermore, a smaller opening of the glottal constriction
leads to larger velocity fluctuations in the supraglottal region, causing a rise in the acoustic contribution of quadrupole sources in the turbulent flow (cf. Sec. 3.3.2). Fig. 5.9 shows the instantaneous flow field around the vocal folds with the peak velocity values occurring in the core of the glottal jet.

5.2.2. Vorticity and Coherent Structures

Vortices are produced during the phonation process already in the intraglottal region during the closing phase of the vocal folds, which then take on a divergent shape, as outlined in Fig. 1.3. The process of intraglottal vortex shedding has been first investigated by Mihaescu et al. (2010) by the use of LES on a simplified glottal geometry. Fig. 5.7 shows the trail of low and high pressure regions as a result of vortex shedding from a glottis of divergent angle of 20°.

Additionally, coherent structures in the supraglottal region are produced through impact of the glottal jet at the solid airway walls and in its shear layers towards the resting fluid. The magnitude of vorticity undergoes a sharp rise at the exit of the glottal constriction, as demonstrated in Fig. 5.8, which shows the y-component of vorticity, perpendicular to the plane section. The coherent structures shed in the supraglottal region are convected downstream and introduce fluctuations even far away from the vocal source, contributing to the overall sound output during phonation. This effect is investigated by comparing point-probe Fourier spectra of velocity fluctuations at different levels of the upper airways. As demonstrated in Fig. 5.10 for the case of phonation of the vowel [a], the organisation of the frequency content in the flow changes considerably along the length of the vocal tract. The dominant frequencies $F_1$ and $F_2$ of the spectral envelope, which distinguish the vowel from other spoken
Figure 5.7. Intraglottal pressure distribution at successive time instances (a)-(c), showing a trail of vortices shed from a glottal constriction of divergent shape.

sounds, start to appear close to the end of the vocal tract. Thus, they are likely to be amplified through constructive interference in the mouth cavity.

5.2.3. Broadband Sound Generation

The generation of the inharmonic, high-frequency component of the voice spectrum is due to flow fluctuations in the intra- and supraglottal region (cf. Sec. 2.1). Furthermore, the shape of the glottal constriction impacts the amplitude of these pressure fluctuations, as shown in Fig. 5.11 for three different divergent glottal angles. A sharp increase of the pressure root-mean-square (RMS) is registered for larger angles above 10°. This emphasises the importance of the closing phase of the glottal cycle for the production of near-field flow structures and the resulting broadband component of the emitted sound. Furthermore, a considerable output of pressure fluctuations occurs in the glottal jet shear layers. The glottal jet shows a vena contracta effect, forcing it on a smaller
Figure 5.8. Perpendicular vorticity component in the coronal plane for the convergent cases of -10° (a), -20° (b), and -40° (c) and divergent cases of 10° (d), 20° (e), and 40° (f) of the glottal angle.

cross-sectional diameter and thereby increasing its mean velocity and shear between the jet and the resting fluid (Triep & Brücker 2010; Matheus & Brücker 2018). This is particularly the case for large convergent glottal angles, which occur during the opening phase of the glottal cycle.

In order to separate the aerodynamic and acoustic disturbances in the near field, spatial, two-dimensional Fourier transformations of the pressure fluctuations are computed. The phase velocity of disturbances is defined as \( c_\varphi = f \lambda = f \cdot 2\pi/k \). Since the wavenumber \( k \) is subsequently plotted in the range \([-\pi, \pi]\], we use \( c_\varphi = f \cdot \pi/k \). Therefore, the group velocity of disturbances can be computed via \( c_g = \pi \cdot \partial f/\partial k \), with the slope of a frequency-wavenumber diagram \( \partial f/\partial k \). Fig. 5.12(a) shows the result for vowel [a]. Both acoustic waves
Figure 5.9. Snapshot of the velocity field at the glottis showing the glottal jet in the sagittal plane section for the airway geometry of the vowel [u]. The position of probes P1, P2, P3, P4, as well as $\Phi_{\text{far-field}}$ for the monitoring of velocity and pressure fluctuations is indicated.

Propagating at the speed of sound (marked by the almost vertical dashed lines), as well as aerodynamic disturbances caused by instabilities in the shear layers of the glottal jet (marked by the dashed line of smaller slope). The convection velocity of these flow structures is about half of the mean velocity of the glottal jet. In the considered case of Fig. 5.12(a) this is congruent with a group velocity of approximately $v_g = 9.8 \, \text{m/s}$. Additionally, Fig. 5.12(b) gives a comparison of the energy content at the fundamental frequency $f_0$, as well as at the formant frequencies $f_1$ and $f_2$ of vowel [u]. It can be seen that the amplitude of broadband sound fluctuations increases towards higher frequencies of about $f \sim O(1000)$. 
Figure 5.10. Semilogarithmic plot of the power spectral density of velocity fluctuations at increasing distance from the glottal source for the vowel [a]. The position of the probes $P_1$, $P_3$, and $P_4$ is given in Fig. 5.9. Helmholtz eigenfrequencies and experimental data have been provided by the Speech Modelling Group at the Department of Mathematics and Systems Analysis, Aalto University.

Figure 5.11. Pressure RMS for the cases of divergent glottal angles of $10^\circ$ (a), $20^\circ$ (b), and $40^\circ$ (c).
Figure 5.12. Frequency-wavenumber spectrum of the near-field pressure fluctuations of the vocal tract configuration of vowel [a] (a). The group velocities of the disturbances are marked by the black dashed lines. Cuts at dominant, discrete frequencies are shown in (b).
CHAPTER 6

Aeroacoustics of a Simplified Soft Palate

Various forms of obstructions can affect the human upper airways in case of breathing disorders such as obstructive sleep apnea (OSA). Sec. 2.3 gives an overview of these conditions of the upper respiratory tract. In addition to their detrimental effect on breathing and oxygenation, these obstructions are usually accompanied by tissue vibrations of different amplitudes leading to the generation of sound (Bohadaña et al. 2014). One such example is the oscillation of airway walls close to the lungs, leading to a characteristic sound referred to as wheezing (Meslier et al. 1995). Another common example is the heavy snoring occurring with OSA patients, which stems from vibrations of the pharyngeal airway and in particular the soft palate (cf. Sec. 1.1).

In the context of this work, the mechanical and acoustic response of a flexible bluff body, which serves as a simplified model of the soft palate in the pharyngeal airway as described in Sec. 1.1, have been evaluated for a range of elastic moduli. The presented results are obtained through the numerical investigation of the flow-induced vibrations of the flexible element in cross-flow at low Reynolds numbers of $Re = 100$–$1000$ by means of fluid-structure interaction simulations that are specified in Sec. 4.1. The aeroacoustics in the near field are assessed by direct computation of the compressible airflow (cf. Sec. 3.2.2). Additionally, the Ffowcs Williams-Hawkins (FWH) acoustic analogy (cf. Sec. 3.3.2) is applied, characterising the acoustic sources and the corresponding sound propagation.

6.1. Flow Instabilities and Vortex Generation

The occurrence of vortical structures and their dominant frequencies in the wake of the deforming structure are investigated. The shedding frequency in the downstream region is connected to the object diameter $d$ and the mean inflow velocity $\bar{u}$ by the Strouhal number $St = f\cdot d/\bar{u}$, which is approximately constant over a wide range of the Reynolds number $Re$ (Norberg 1994). Examples of mean of pressure and pressure fluctuations for the considered geometry are given in Fig. 6.1–6.2, showing the regular trail of low pressure cores of vortices convected from the structure.
Figure 6.1. Time-averaged pressure field around the elastic element at a normalised length of $l^* = l/d = 1.75$ and at a Reynolds number of $Re = 300$.

Figure 6.2. Pressure fluctuations at normalised length $l^* = l/d = 1.75$ and Reynolds number $Re = 300$ in the vicinity of the elastic element.

6.1.1. Influence of Reynolds Number

An increase of $Re$ causes a higher number of vortices being shed and increases the lowest frequency registered in the power spectral density of pressure, as shown in Fig. 6.3. Furthermore, upper harmonics are appearing in the cases of higher Reynolds numbers. These are due to the interaction of coherent structures in the main trail of vortices. However, turbulence is reached in the considered cases only at Reynolds numbers well above $Re \sim O(1000)$.

6.1.2. Effect of Appendage Length

The considered case is investigated for different non-dimensional lengths $l^* = l/d$ and Reynolds numbers $Re$ at a constant elastic modulus $E$. Tab. 6.1 shows the resulting values of the dominant shedding frequencies of an element of $E = 0.1$ MPa in terms of $St$. The data is obtained through Fourier analysis based on point probe measurement of the pressure close to the object. The position of the probe is indicated in Fig. 6.3.

An increase of appendage length leads to a decrease of the vortex shedding frequency. Additionally, laminarisation of the flow takes place at a particular
Figure 6.3. The effect of increase of Reynolds number $Re$ on the flow structures in terms of the normalised perpendicular vorticity component is shown in (a)-(e). The frequency content of the wake flow is given in terms of the power spectral density (PSD) of pressure for Strouhal number $St$.

length, at which the vortex shedding is inhibited completely. At these lengths, the otherwise unsteady wake flow becomes steady and laminar again. This critical length varies and increases with Reynolds number.

6.2. Vibrations and Dominant Frequencies

The pressure fluctuations in the wake region of the solid obstacle are composed of two parts: Aerodynamic pressure unsteadiness due to shed vorticity being convected downstream on the one hand, and the acoustic pressure fluctuations generated by moving and stationary surfaces on the other hand. At conditions
Table 6.1. Dominant shedding frequencies for an elastic element of Young’s modulus $E = 0.1$ MPa terms of the Strouhal number $St$. The cases of steady, laminar flow in the domain are marked by a dash - , while cases with mainly broadband fluctuations are assigned by a star *.

<table>
<thead>
<tr>
<th>Normalised length</th>
<th>$Re = 100$</th>
<th>$Re = 150$</th>
<th>$Re = 300$</th>
<th>$Re = 500$</th>
<th>$Re = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>without</td>
<td>0.30</td>
<td>0.31</td>
<td>0.33</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>0.875</td>
<td>0.32</td>
<td>0.32</td>
<td>0.35</td>
<td>0.18*</td>
<td>*</td>
</tr>
<tr>
<td>1.75</td>
<td>-</td>
<td>0.36</td>
<td>0.35</td>
<td>0.34</td>
<td>0.28</td>
</tr>
<tr>
<td>2.625</td>
<td>-</td>
<td>-</td>
<td>0.27</td>
<td>0.27</td>
<td>*</td>
</tr>
<tr>
<td>3.5</td>
<td>-</td>
<td>-</td>
<td>0.24*</td>
<td>0.22*</td>
<td>*</td>
</tr>
</tbody>
</table>

close to resonance, vibrations and displacement of elastic parts are expected to considerably increase the emission of acoustic frequencies.

6.2.1. Effect of Elasticity

The displacement of the flexible element is larger for higher Reynolds numbers. Additionally, it is found that a decrease of Young’s modulus causes both a significant increase of the amplitude of oscillations as well as an appearance of higher order modes in the frequency content of the vibrations. As the elasticity is reduced, the Strouhal number of the higher mode in the spectrum drops in proportion and its peak is shifted towards the lower end of the spectrum.

Elastic structures subjected to unsteady flow of low Reynolds number tend to exhibit strong oscillatory behaviour, if the frequency of the shed vorticity in the flow lies in the vicinity of the natural frequency of the structure. The amplitude of the vibrational response then depends heavily on the elastic modulus. In extreme cases resonance can be obtained and the magnitude of the displacements reaches a maximum.

6.2.2. Resonance

If the frequency of impinging coherent structures and one of the eigenfrequencies of the structure are matching, resonance occurs. This behaviour is investigated under variation of elasticity and Reynolds number with the results depicted in Fig. 6.4. It can be recognised that the oscillations of second dominant modes at higher stiffness of the element show consistently higher frequency and Strouhal number for all investigated Reynolds numbers (cf. Fig. 6.4(a)-(b)). In these cases resonance is not reached. However, as the stiffness is reduced, the curve of second dominant modes is approaching the limit of the flow-induced vortex shedding frequency. Finally, resonance is observed at the point of intersection of the two curves (cf. Fig. 6.4(c)-(d)). The Strouhal
number of the periodic vortex production is approximately constant over the considered range of Reynolds numbers, as demonstrated in Tab. 6.1.

![Graphs](image)

**Figure 6.4.** Resonance of the structure as shown for low Young’s modulus and high Reynolds number at the matching of first and second dominant mode in the frequency spectrum of structural oscillations. Plots (a)-(d) show the occurrence of modes in the Fourier transform of the displacement for materials of elasticity $E = 1\text{ MPa}$ to $E = 0.01\text{ MPa}$. The first dominant modes relate to the frequency of flow-induced vortex production. The second dominant modes represent higher order vibrational frequencies dependent on the stiffness of the structure.
CHAPTER 7

Conclusions and Outlook

The presented work aims at modelling the physical processes involved in the production and propagation of sound in the human upper airways by means of numerical methods. Due to the airways being a complex system of channels, experimental investigation of the process of speech, as well as the underlying mechanisms of obstructions remain challenging. Numerical flow simulations applied to realistic models of the vocal and respiratory tract have therefore taken a promising role due to the increase in computational power and the non-invasiveness.

Sound Through Voice Production

An initial study based on reduced-order acoustic models has been carried out. The consequences of changes of the waveform of the glottal pulse on the articulated sound, as they might occur as a result of upper airway diseases, have been evaluated, leading to the following findings (cf. Paper 1):

- A decrease of the vocal tract length leads to an increase in both $F_1$ and $F_2$ formant modes and thus to a shift of the vowel points in the frequency space (i.e. formant map) to the upper right and thus to higher frequencies.
- An increase of the relative glottal opening fraction leads to a slight drop of the dominant vowel frequencies. The glottal opening quotient is higher in cases of lesions and polyps at the vocal folds, or for disorders leading to reduced tissue elasticity and incomplete glottal closure.

The chosen approach of direct computation of articulated vowel sounds via compressible flow simulations using large eddy simulations (LES) and a glottal waveform model has been applied on realistic, static vocal tract models of different vowel pronunciations of a healthy male subject, which were obtained through magnetic resonance imaging (MRI) (cf. Paper 2-3):

- The computational method using LES is validated by comparison of the dominant modes in the far-field acoustic spectrum with Helmholtz resonance frequencies of the vocal tract and results from experimental studies. An initial verification study based on a rigid glottal model yields
surface pressure data in good agreement with published experimental results.

- Vorticity and velocity fields were monitored at the level of the vocal folds, capturing the coherent structures produced at the glottal constriction.
- Fourier analysis of the velocity fluctuations along the length of the vocal tract have been carried out, indicating the approximate region of amplification of the formant frequencies.
- Fourier surface spectra reveal the cavity-association of the first two formant modes of the considered vowel geometries. The critical locations in the upper airways leading to constructive and destructive interference of acoustic waves, which are exciting the resonance modes of the airway under voiced speech, could be identified.

The quasi-steady approximation in conjunction with LES is applied to a realistic vowel geometry to assess the acoustic sources in the supraglottal near field (cf. Paper 4):

- The dominant frequencies of the inharmonic, aerodynamic sound generation are found through cross-correlations between near- and far-field histories of velocity and pressure fluctuations.
- The acoustic sources in the vicinity of the glottal jet and the supraglottal region are localised through Lighthill’s acoustic analogy for various convergent and divergent glottal shapes.

**Sound Through Airway Obstructions**

An idealised model of the pharyngeal channel with a flexible structure resembling the human soft palate was considered to study fluid-structure-acoustics interactions. The impact of critical parameters of the flow and structure on the onset of vibrations and the dominant frequencies in the near field was explored, leading to several conclusions (cf. Paper 5):

- Larger amplitudes of the structural oscillations are observed for a decrease in $E$ and increases in $l^*$ and $Re$. By artificially increasing the elastic modulus, the tissue of the soft palate could be prohibited from resonating or collapsing. This concept is already in use through the so-called pillar procedure, a minimally invasive method that involves the placement of thin polyester rods in the soft palate.
- In the case of lower Young’s moduli additional superimposed oscillation frequencies are observed, which could couple with the flow to a resonance effect. This is likely to occur for structures of $E < 0.05$ MPa as the flow approaches $Re = 1000$.
- It is shown that there exists a critical element length in the range of lower Reynolds numbers, for which the unsteadiness in the wake region is suppressed and vortex shedding does not occur.
• The computed pressure fluctuation field near the flexible element shows perturbations in the acoustic near field similar to a dipole field.
• The dominating sound source of the investigated structure changes from dipole at low Re and high E to monopole at high Re and low E. This is due to the strong vibrational response at low stiffness causing an increase in emitted sound power of the monopole. The frequency of sound waves emitted by the monopole lies consistently in the audible range above 100 Hz.

Scientific Contributions
The following contributions have been made through this work:
• A method is introduced to resolve both harmonic and inharmonic component of a voiced speech signal by using realistic upper airway geometries obtained through MRI and compressible flow simulations (cf. Paper 1-2).
• The flow structures occurring during voice generation are characterised in terms of their group velocity and energy content (cf. Paper 3).
• The dominant frequencies of flow-induced sound in the near field are determined for various glottal shapes by application of acoustic analogies (cf. Paper 4).
• Conditions of excitation and resonance of a simple elastic structure of physiological properties are established (cf. Paper 5).

Challenges and Limitations
Due to the inherent complexity of sound production and propagation several assumptions and simplifying modelling approaches have been taken. In the following, the most important limitations of this work are outlined.
• In the current study the glottal excitation is imposed through an analytic model of the volumetric flow rate, a glottal waveform. This was necessary for several reasons: First, a fully coupled fluid-structure interaction between vocal folds and airflow led to an improper source signal. Secondly, the contact problem made full closure of the glottis impossible during numerical flow simulations, or led to negative volume cells. Thirdly, the mesh deformation that occurs both with imposed vocal fold motion, as well as with excited vibrations, led to a large uncertainty of the acoustic wave propagation by the LES simulations of the compressible flow.
• The underlying work considers the upper airways as static. While it can be argued that flexible walls barely influence the propagation of acoustic modes, a complete and physically accurate model of the upper airways requires the airway to be elastic to predict the impact of wall vibrations on far-field acoustics.
• In the current approach, the airflow is treated as isothermal. Real airflow through the human upper airways is in fact warm and rather moist, introducing slight variations in the local speed of sound c and gas constant R. Although this simplifying assumption worked well for the considered acoustical data, a more thorough treatment would entail temperature variations.

Future Work

This thesis represents a report on the findings of this doctoral project. Therefore, due to restrictions in time, many important questions surrounding the topic are still to be answered. In the following, possible improvements and further objectives are given.

• Despite the above challenges, the construction of a stable, two-way coupled model of the vocal folds in conjunction with a realistic vocal tract should be the ultimate goal of computational voice acoustics. This could be achieved through coupled application of finite element and finite volume solvers.

• For the reestablishment of speech in case of vocal disorders and a targeted therapy, the identification of critical parameters of the source signal and the upper airways influencing the speech signal is desired. By varying parameters of the glottal waveform and the airway geometry, their impact on the formant positions, voice loudness, and intelligibility could be assessed.

• Correlations between anatomical and mechanical properties of the upper airways and airway collapse and sound production should be examined for more complex cases of obstructions. The analysis will allow making associations between the sound produced and the severity of the obstruction.
Acknowledgements

This project was supported by the Swedish Research Council (Vetenskapsrådet) through the grant VR 621-2012-4256. This contribution is gratefully acknowledged.

Computational resources used for the simulations in this work have been provided by the Swedish National Infrastructure for Computing (SNIC) and the Parallel Computing Center (PDC) at KTH.

The Erik Petersohns Minne foundation is gratefully acknowledged for the award of two travel scholarships.

Furthermore, I would like to express my deep gratitude to my supervisors Mihai Mihaescu and Anders Dahlkild for accepting me and trusting me with this interesting project. Thank you Mihai for letting me pursue my own ideas, providing guidance whenever I needed it, and for showing understanding for the challenges. Thank you Anders for sharing your valuable knowledge in fluid mechanics and thermodynamics with an always open door.

Thanks go also to the other members of the research group Laszlo Fuchs, Lisa Prahl-Wittberg, Christophe Duwig, and Mireia Altimira for their much appreciated comments both during and after the Monday morning seminars and regular group meetings.

Thanks to Shervin Bagheri for reviewing this thesis and providing helpful comments.

Jarno Malinen and the Speech Modelling Group at Aalto University in Finland are acknowledged for providing the MRI data used for the airway geometries in this study and discussing the obtained results, helping me understand speech acoustics much better.

The participants of the InnerVoice group, hosted by the Division of Speech, Music and Hearing at KTH are acknowledged for fruitful conversations and seminars.

I am thankful to have been surrounded by such great colleagues as Bernhard, Alexander, Timea, Tomas, Elias, Shyang-Maw, Bertrand, Niclas, Asuka, Song,
Alex, Emelie, Francesco, and Roberto. Thanks for your great company, our conversations, both on- and off-topic, and the countless lunches we had together.

Cheers to the Imperial College alumni group and Tom, Scott, Mark, Simon, Kieran, and Malcolm for keeping in touch over the years and for the travel adventures. Particular thanks go to Paulina, Willem, and Romain for our shared experiences here in Stockholm. I am glad that I have met you.

Vielen Dank an meine Familie für die bedingungslose Unterstützung, dass Ihr mich auf diesen Weg gesetzt habt und mich zurück auf die richtige Bahn holt, wenn es nötig ist.

January 2019, Stockholm

Lukas Schickhofer
References


BLAUSEN MEDICAL GALLERY 2014 Wixiversity Journal of Medicine, , vol. 1.


ERATH, B. D. & PLESNIAK, M. W. 2012 Three-dimensional laryngeal flow fields


Howe, M. & McGowan, R. 2007 Sound generated by aerodynamic sources near a


Liljencrentris, J. 1985b Speech synthesis with a reflection-type line analog, , vol. 85. Royal Institute of Technology.


Löhner, R., Cemral, J., Soto, O., Yim, P. & Burgess, J. E. 2003 Applications


Mathieson, L. 2013 *Greene and Mathieson’s The Voice and its Disorders.* John Wiley & Sons.


McNicholas, W., ed. 2013 *New Standards and Guidelines for Drivers with Obstructive Sleep Apnoea Syndrome,* Obstructive Sleep Apnoea Working Group.


Naughton, M. T. 2006 The link between obstructive sleep apnea and heart failure: underappreciated opportunity for treatment. *Current Heart Failure Reports* **3** (4), 183–188.


SIMCENTER STAR-CCM+ USER GUIDE 2018 Version 13.04.011.


STORY, B. H. 2015 3 mechanisms of voice production. The Handbook of Speech Production p. 34.


Zhao, M., Barber, T., Cistulli, P., Sutherland, K. & Rosengarten, G. 2013a Computational fluid dynamics for the assessment of upper airway response to oral appliance treatment in obstructive sleep apnea. Journal of Biomechanics 46 (1), 142–150.


Part II

Papers
Summary of Papers and Author Contributions


Summary: Vocal tract acoustics under healthy, voiced speech are investigated by applying a one-dimensional wave reflection analogue, as well as acoustic pressure data from large eddy simulations. A parameter study is performed to investigate the effect of variations of the fundamental frequency, the glottal opening quotient, and the vocal tract length on the first two formant frequencies.

Contributions: All simulations were run by LS. The interpretation of the results was discussed by all authors. The manuscript was written by LS with input from AD and MM.


Summary: Large eddy simulations are performed for direct aeroacoustics calculations of a realistic vocal tract model. Boundary conditions, spatial and temporal resolutions were tested and first data of vorticity and $\lambda_3$-criterion, as well as pressure fluctuations, were computed. This study was meant as a precursor leading up to the more extensive study presented in Paper 3.

Contributions: A realistic upper airway geometry, provided by the Speech Modelling Group at the Department of Mathematics and Systems Analysis, Aalto University, was implemented by LS to study the vocal tract acoustics under voiced speech. LS created the computational grid, tested appropriate boundary conditions, and performed the simulations. The manuscript was written by LS with feedback from AD and MM.


Summary: A method for the computation of both harmonic and inharmonic
sound components of phonation is presented based on realistic vocal tract geometries of various vowel pronunciations obtained through magnetic resonance imaging (MRI). The presented approach is validated and verified through comparison with experimental data from surface pressure measurements and sound recordings. Acoustic and convected disturbances are identified in the near and far field through frequency-wavenumber spectra. Moreover, the origins of formant modes of different universal vowels are located and their cavity-association obtained through surface sound pressure levels.

Contributions: LS set up the cases and performed the computations and post-processing. The results were discussed by all authors throughout the study. The manuscript was written by LS, with exception of Sec. 2.2 and parts of Sec. 2.1, which were written by JM.

Paper 4: L. Schickhofer, M. Mihaescu. Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes. To be submitted.

Summary: The quasi-steady approximation of aeroacoustics is used to study the aerodynamic sound generation during different instances of the glottal cycle. Thus, the vocal tract geometry of vowel [a] is modified in the glottal domain to account for convergent and divergent vocal fold shapes. The flow field is investigated and the fluctuations of the flow variables are computed for the supraglottal region. The energy contained in near-field fluctuations is found to increase towards higher divergent angles. Increases in shear-layer instabilities and convected disturbances are detected for the convergent shapes due to a vena contracta effect. Furthermore, acoustic sources are localised in the coronal and sagittal plane sections through the epilarynx by applying Lighthill’s acoustic analogy and the Lamb vector divergence. An increase in strength of the sources is noticed both towards higher convergent and divergent angles.

Contributions: LS prepared the geometries, ran the simulations, and evaluated the results. LS prepared the manuscript with feedback from MM.


Summary: The conditions for flow-induced vibrations in the upper airways have been assessed through a parameter study with numerical simulations of an idealised geometry in compressible flow. The dominant frequencies of structural vibrations alongside the frequencies of vortex shedding in the near field of the elastic element are calculated. An acoustic analogy based on the Flowcs Williams-Hawkings (FWH) equation solved for confined channel flow is applied to compute the acoustic sources and their contribution to the overall pressure fluctuations.
Contributions: LS set up the case and ran the simulations under supervision by MM. LS carried out the post-processing along with the writing of the manuscript. MM gave advice on the use of acoustic analogies. AD and MM provided feedback throughout all stages.

Other publications:
The following manuscript has been published in the framework of this thesis, but is not included.
